12-1-1996

Converting Publications in the Air Force to SGML: A Recommendation Based on Commercial Experience

Shari T. Miles

CONVERTING PUBLICATIONS IN THE AIR FORCE TO SGML: A RECOMMENDATION BASED ON COMMERCIAL EXPERIENCE

THESIS

Shari T. Miles, Captain, USAF

AFIT/GIR/LAR/96D-5

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GIR/LAR/96D-5

CONVERTING PUBLICATIONS IN THE AIR FORCE TO SGML: A RECOMMENDATION BASED ON COMMERCIAL EXPERIENCE

THESIS

Shari T. Miles, Captain, USAF

AFIT/GIR/LAR/96D-5

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

AFIT/GIR/LAR/96D-5

CONVERTING PUBLICATIONS IN THE AIR FORCE TO SGML: A RECOMMENDATION BASED ON COMMERCIAL EXPERIENCE

THESIS

Presented to the Faculty of the Graduate School of

Logistics and Acquisition Management of the

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Information Resource Management

Shari T. Miles

Captain, USAF

December 1996

Acknowledgments

I would like to thank my thesis advisor, Dr. Freda Stohrer, for her expertise, guidance, and support. Her efforts, along with the attention to detail and excellent inputs provided by my thesis reader, Dr. Alan Heminger, made this research project a success. They were able to provide me with the insight and perspective which helped me focus on my research topic. I would not have successfully made it without their knowledge and guidance during this lengthy process.

My most sincere appreciation goes to my husband, Rick, for his continuous encouragement and support throughout this AFIT experience. The long distance telephone calls and weekend trips were well worth the effort.

Shari T. Miles

Table of Contents

Page

Page

List of Figures

List of Tables

Page

Abstract

The objective of this research was to determine those benefits of SGML that the Air Force can use in its publications process and to determine if benchmarking can improve the Air Force's migration to an electronic publishing environment. To accomplish this, the author completed an exploratory, qualitative study which defined the components and benefits of SGML and the benchmarking application and process. The qualitative techniques for this research included a literature review and questionnaire. The literature review covered both SGML and benchmarking. The questionnaire was an exploratory instrument that was sent to 60 commercial enterprises.

The results of the questionnaire confirmed and elaborated the findings of the literature review. In general, this analysis indicates that the Air Force electronic publishing environment can benefit from the use of SGML.

SGML allows for the creation of a document that can be reused in many ways and allows for search, retrieval, and transfer of data across networks and platforms. It was also determined that there are many capabilities and benefits associated with the use of SGML and there are ways in which the Air Force publishing centers can use the full potential of SGML. The five most cited benefits of SGML are machine independence, reusability, shareability, increased productivity, and improved data integrity.

Last, the analysis indicated that a benchmark of commercial enterprises would provide useful information in streamlining the Air Force's conversion to SGML.

Benchmarking provides an opportunity to examine the best practices of commercial enterprises and then implement changes based on the observations.

CONVERTING PUBLICATIONS IN THE AIR FORCE TO SGML: A RECOMMENDATION BASED ON COMMERCIAL EXPERIENCE

I. Introduction

Background

The Department of Defense (DoD) is facing a period of profound change in the country's national defense strategy. The defense budget is 40 percent lower than in 1985, the peak year for defense spending, and there are 33 percent fewer military personnel than at peak force size in the mid 1980's (White, 1996; DoD, 1996: WWW). These reductions have created a new agenda for how DoD does business. The new agenda, in turn, is driving DoD to change the way it uses information technology.

As the armed services have restructured, they have had to look at how to perform their day-to-day business activities more efficiently and effectively. The Services have consolidated their major commands and reorganized their base structure to accomplish new missions while working with reductions in personnel and money. Many of the day-to-day functions performed by individuals or organizations have been automated uniquely by each Service. This automation has served to streamline the activities within the Service. This is as true of DoD information systems as for its other systems. However, these unique stand-alone information systems are not always compatible and do not necessarily interface with related information systems within any one Service or across Services. Incompatibilities exist with operating systems and application software, input

and output devices, applications, and languages. The incompatibilities greatly reduce the effectiveness and efficiency of the affected processes. These inefficiencies within and across Services have led to the need for one standardized approach which can eliminate the duplication and redundancy of information within business functions and automated systems. Without standards, the use of the information within the automated systems is limited to those who possess the software applications to access that information.

These changes within DoD have forced the Air Force to analyze how its organizations are managed and how work is performed, and then to implement changes that will make its people more productive. During a speech to the House National Security Committee, Sheila Widnall, Secretary of the Air Force, said "we are 'modernizing our users' (military members) to make faster and better use of information. These are precisely the advances we need to fully exploit the capabilities of a much smaller military" (Widnall, 1995). Such advances can be accomplished only if new technologies available through the information revolution are applied to information management practices. Officials within the Air Force publishing environment are addressing how they can improve their management practices.

According to the Air Force Electronic Publishing Master Program Guide (1996), the Air Force plans to migrate from paper to electronic publishing. By replacing the traditional paper-based publishing system with an electronic publishing environment, the Air Force will be able to realize the value of sharing and reuse of information amongst organizations whether they are located together on one base or spread across various remote locations.

Since the Air Force came into being, administrative information was distributed through a paper-based system. Any department, from the Secretary of the Air Force to a two-letter office at base level, can request that a directive, instruction, pamphlet, or other type of document be published. Developing publications requires many levels of coordination and each publication must be in the proper format and style of language as prescribed by AFI 37-160, Volume 1. Once the document is published and officially a new document, a publishing bulletin is transmitted to every publication representative from the headquarters level to base level announcing the new publication. If the publication is necessary to support an organization's mission, the organization's publication representative requests it. The publication will arrive in paper copy or on a CD-ROM through the base mail system. Often, the CD-ROM version requires the user to convert it to a format that can be used on the software application available on the user's computer system before the user can access or print the requested document. Also, if the publication has a supplement, the publication representative is required to post the supplement reference to the original paper document manually and annotate the date/time next to each entry. This process can be time consuming and labor intensive if the supplement has made major revisions to the original document. The publication representative is also responsible for maintaining an adequate supply of publications to meet the needs of his/her organization. During wartime, requirements direct a 30 to 60-day supply of forms for day-to-day operations and a 90-day supply of necessary publications to support the War Mobilization Plan and Emergency Mobilization Plan.

Although the paper-based system works as intended, today's business environment with its decreasing manpower, increasing costs, and mounting expectations force changes in the publishing process. An electronic system for the publication production and distribution process would improve its efficiency and effectiveness. A standardized approach to the electronic publishing system would also allow the Air Force to transmit publications across varying machines and software applications without the need to re-key the data or filter it in order to use it again. This non-proprietary format allows the data to be used and reused in any way by any software application. An electronic system provides a publishing library in a central repository for the publication representative to access at any time. Also, the posting of supplements can be integrated automatically, thereby eliminating the need for someone to post the changes manually. This integration automatically links the publication and its supplement and allows the end user to see the two documents as one document. Last but not least, the electronic environment would assist the publication representative in maintaining the required supply of documents and forms by providing him/her with access to whatever is needed at any time. When the publication representative needs a document he/she can access it at any time and print out whatever is required. Thus, access to a central repository of publications eliminates the need to wait for the document to arrive in the base mail system. The information requested gets to the right people in the time required. The electronic system also eliminates the need for warehouse space to hold the required supply of documents. There is no need to rely on the paper-based system if our hardware

and software applications can read all publishing documents and deliver the information to the right people at the right time.

There are many ways the Air Force can accomplish its move to electronic publishing. Some options include putting publications on the Internet and World Wide Web (WWW) for users to have direct on-line access, contracting with a commercial publishing company to convert the paper-based publications and create direct on-line delivery service for users, or leasing an on-line service such as CompuServe to provide direct access to a central repository of publications. There are advantages and disadvantages of using on-line services. The advantage of on-line services is that they provide the user with much more data by being served from a central repository. Another advantage of an on-line service is the immediacy of the data (Travis, 1995: 55). Users will always get up-to-date, timely information since they always have access to the repository. The main disadvantage of on-line services is the cost (Travis, 1995: 55). An on-line service, such as CompuServe, costs money every time the user accesses the network. Military members are constantly accessing publications in the performance of their day-to-day tasks and a fee to access the necessary information would not be practical. The arrival of the WWW has made access to information easy and simple. Documents can be placed at a site accessible by users of the WWW via a local telephone (Travis, 1995: 55).

The Air Force's conversion to electronic publishing encompasses the current publishing business from Headquarters Air Force to base level, from the author to end user. The Air Force publishing environment includes over 13,000 paper products in the

form of such documents as Air Force Instructions, Pamphlets, Handbooks, and Manuals (see figure 1).

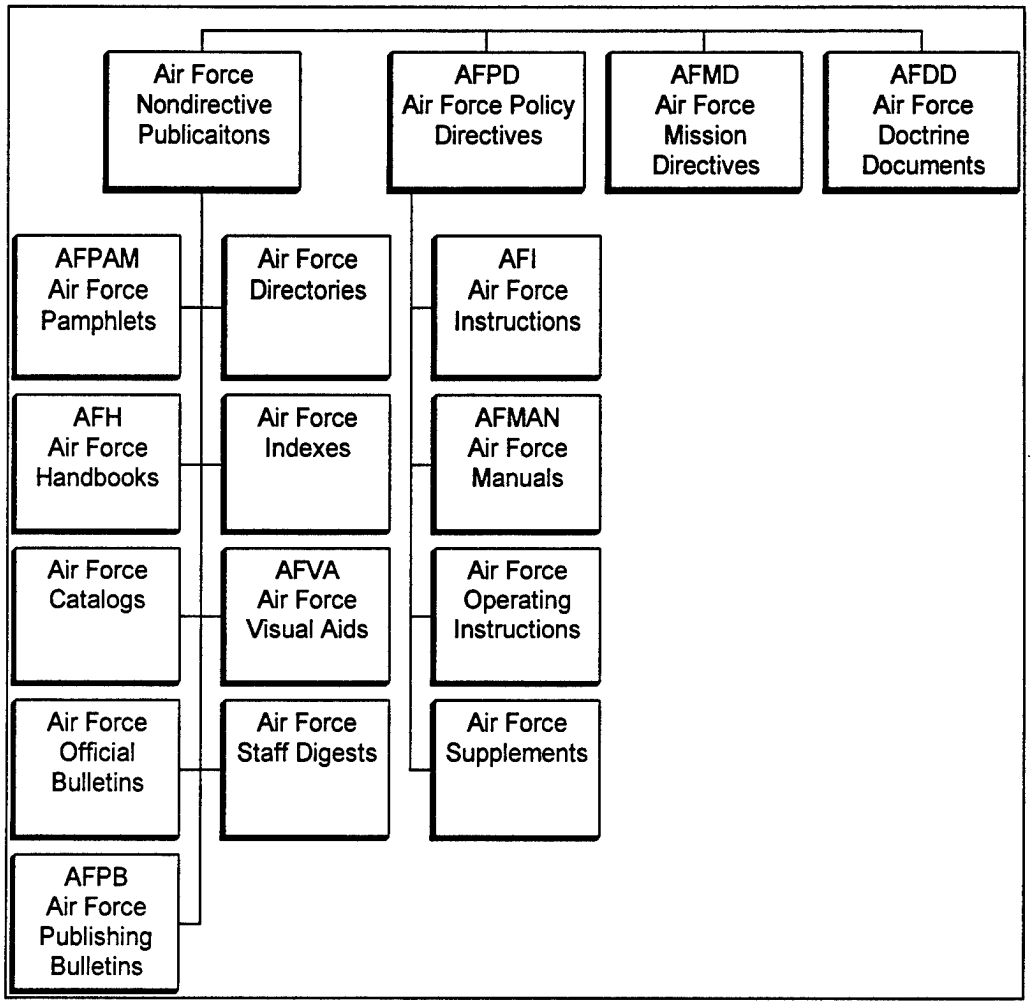

Figure 1: Authorized Air Force Publication Types (AFI 37-160, Vol 1: 7)

The Air Force adopted the Standard Generalized Markup Language (SGML) as a tool to meet the Continuous Acquisition Life-Cycle Support (CALS) mission in 1986. SGML is a vendor-independent documentation interchange language used for structuring the content of documents. SGML describes the document's structure independent of any

particular computer operating system and application software, input and output devices, applications, or languages (Heimburger 1994: 241). System independence means that document structure is not tied to one specific system and its use should be standard on any manufacturer's system (Heimburger, 1994: 241). SGML achieves system-independence by preserving the content (information) and the structure (relationship of the data) of documents regardless of when the document was created (November 1995, WWW). Device independence allows an individual within the organization to use any type of hardware for its documents without being tied to a specific piece of equipment. Application independence allows the standard, SGML, to be used in describing the structure for simple or complex documents (Heimburger, 1944: 242). SGML's language independence allows it to be used equally well in different countries.

Since initial adoption for the CALS mission, SGML has proved to be a powerful data interchange standard that will meet the needs of the Air Force electronic publishing program to improve efficiency and effectiveness of the publication process (USAF Master Program Guide, 1996). For this reason, SGML was adopted as the means for migrating the Air Force from a paper-based environment toward an electronic user environment (USAF Master Program Guide, 1996). SGML gives an organization the capability to publish several different documents of the same type or the same document in several formats. In an SGML document, format is separated from content by the use of tags. Tags consists of pieces of code surrounded by the symbols "<" and ">" and are used to indicate structural information or formatting. Tags are interspersed with the text of the document in order to identify the type of content within the document structure.

These tags mark the beginning and end of each part within the overall structure. When format is separated from content, which is the case when using SGML, the document can be reused in many ways. SGML also allows for search, retrieval, and transfer of data across networks and platforms (open systems architecture). These benefits provided by the use of SGML offers capabilities that are not available in the paper-based environment. These benefits will be discussed in Chapter Two.

Problem Statement

This thesis will determine those benefits of SGML that can be used by the Air Force to enhance future publications and determine if benchmarking can improve the Air Force's migration to an electronic publishing environment.

Research Questions

The specific research questions are:

1. What are the benefits of using SGML that will be of value to the Air Force?
2. Which commercial enterprises are using SGML?
3. Will those commercial enterprises identified in research question two be able and willing to provide information regarding the benefits of using SGML? If so, what are the benefits of using SGML in the commercial enterprise?
4. What underlying problems or concerns are associated with these commercial enterprises achieving the benefits of SGML?
5. Does SGML training affect an organization's successful use of SGML?
6. Can a commercial enterprise be used as a benchmark for the Air Force in its migration to an electronic publishing environment?

Need for Research

DoD operations are undergoing change and the Air Force must also make changes if it wants to successfully manage and operate its business processes throughout the Service. The Air Force is working toward the use of common, shared data, and to accomplish this a complete understanding of SGML must be reached to ensure the Air Force's implementation to a fully electronic publication environment is successful. In essence, this thesis will provide information on SGML's components, benefits, and predict the impact of the conversion of Air Force publications based on the experience of commercial enterprises that have been using SGML. Lastly, the research will provide an evaluation of benchmarking concepts and applications to determine if a commercial enterprise can be used as a benchmark for the Air Force.

Definition of Terms

A glossary of terms used in this thesis may be found in Appendix A.

Summary

Chapter One provides background to the problem, presents a statement of the problem to be researched, the major research questions, and significance of the research. Chapter Two provides a literature review of SGML's components and benefits and also presents a discussion on benchmarking concepts and applications. The methodology used to investigate the research questions outlined in this chapter will be described in Chapter Three, followed by Chapter Four which will present the findings of the

investigative questions and observations. Finally, Chapter Five presents a discussion of the results, along with recommendations for further research.

II. Literature Review

Introduction

According to the Air Force Electronic Publishing Master Program Guide (1996), the Air Force plans to migrate from a paper-based publishing environment to a standardized electronic publishing environment. The Air Force Electronic Publishing Program, which began in 1993, has proposed the conversion of all Air Force paper products in the publishing inventory to a standard electronic product. Since the initiation of this program, the Air Force has converted many of its publications from paper-based publishing to a more advanced, search capable CD-ROM format which provides on-line access to publications. The Electronic Publishing Program has proposed the integration of SGML into the publication process.

There are a number of ways the Air Force can accomplish its move to electronic publishing. Some options include putting publications on the Internet and World Wide Web (WWW) for users to have direct on-line access, contracting with a commercial publishing company to convert the paper-based publications and create direct on-line delivery service for users, or accessing an on-line service such as CompuServe to have direct access to a central repository of publications. Currently, publications are developed and requested at all levels within the service. Therefore, the Air Force needs an approach that can manage the electronic publishing process from Headquarters Air Force to base level, from the author to end user.

The Air Force adopted the Standard Generalized Markup Language (SGML) as a tool to meet the Continuous Acquisition Life-Cycle Support (CALS) mission in 1986. Initial evidence suggests that SGML has proved to be a powerful data interchange standard that will meet the needs of the Air Force electronic publishing program to improve efficiency and effectiveness of the publication process. It is for this reason that SGML was adopted as the means for migrating Air Force paper-based published products toward an electronic user environment. According to the Air Force Electronic Publishing Master Program Guide (1996), all publishing activities throughout the Air Force are now using computers to produce paper publications, so the change from paper to electronic publishing is within reach.

With the increased use of computers in the Air Force, publication managers are relying on text processing and other computer-based applications for the publications they develop and maintain. However, the format in which the documents are saved in word processing software is often incompatible between software applications and across hardware platforms. For example, WordPerfect cannot read a document that was created and saved in Microsoft Word unless the document was first converted to a WordPerfect acceptable format. This incompatibility requires documents to be converted to a format that can be accessed by any software. Thus, to transfer information electronically between computers worldwide, document structures must be defined in a standardized way (Heimburger, 1994: 240).

In addition, publications must be kept current to provide up-to-date information to users. Filing manual changes to each individual publication can be tedious and error

prone. Because the filing of manual changes is often given low priority, users may find outdated, incorrect information that could lead to operational errors (Reynolds, 1992: 53).

The Air Force move to change from paper to electronic publishing was initiated in 1993; however, to exploit SGML's benefits fully, the publication centers need thorough research in the area of SGML (USAF Master Program Guide, 1996). Therefore, the analysis of the literature will focus on the potential of SGML, how to exploit the features of SGML, and review commercial enterprises that have or are currently using SGML. A literature review of benchmarking is also presented. The analysis of the information gathered on benchmarking will determine if benchmarking concepts and applications can improve and enhance the Air Force's migration to an electronic publishing environment.

Basic Concepts of SGML

What is SGML? SGML stands for Standard Generalized Markup Language. Let's look at each word in SGML to see what it adds to the concept.

Standard - SGML is recognized as an international standard, ISO 8879, so that communication can occur internationally (Smith, 1992: 14).

Generalized - This term refers to the aspects of SGML. SGML allows documents to be marked up and stored in a neutral format (Heimburger, 1994: 241 and Smith, 1992: 14). This neutrality is achieved with the use of generic identifiers instead of application specific identifiers. SGML is also said to be generalized because it is independent of systems, devices, applications, and languages.

Markup - Markup is defined as text added to the data of a document in order to convey information about it and to indicate its structure (Heimburger, 1994: 243). SGML states how to mark up a document and specifies rules which define how markup is added to the document.

Language - SGML is not really a language. It is a metalanguage; it has syntax but not semantics (Heimburger, 1994: 241 and Smith, 1992: 15). According to Smith, SGML is a representation language, and can be used to describe any document (Smith, 1994: 15).

A typical document can be broken into three layers: structure, content, and style (Heimburger, 1994: 242; Wright, 1992: 280; and Gilmore, 1993: 213). SGML separates these three layers, but deals mainly with the relationship between structure and content. The best way to understand how SGML works is to use an example. Figure 2 shows us a simple memo as an example.

To: Students of SGML
From: Ms. Gilmore
Date: 23 January 1996
Subject: Document Type Definition (DTD) for a Memo

A good way to learn about SGML is to use it, beginning with simple documents.

I am sending you a copy of the Memo DTD. Please note that the following structural rules are built into the memo by the element definitions (called "declarations" in SGML):

- There can be multiple recipients and authors.
- There can be only one date
- Providing a subject for the memo is optional
- A body must follow a header.
- A body must contain at least one paragraph
- Paragraphs can contain one or more list items but these are optional.

I hope that you enjoy studying and using this DTD.

Figure 2: Document Memo (Gilmore, 1993: 213)

Structure: The Document Type Definition (DTD) describes the structure of a document. The DTD functions as a template to identify the type of document -- memo, regulation, official letter, operating instruction -- and the elements of the document in the order they appear (November 1995, WWW). The DTD can be general enough so that it applies to all documents of the same type. In other words, if structured properly, one DTD can serve for most documents that are of the same type. There is no need to recreate a DTD each time the user writes a memo. The original DTD for the memo can

be used over and over again. Figure 3 shows the structure of the document memo pictured in figure 2.

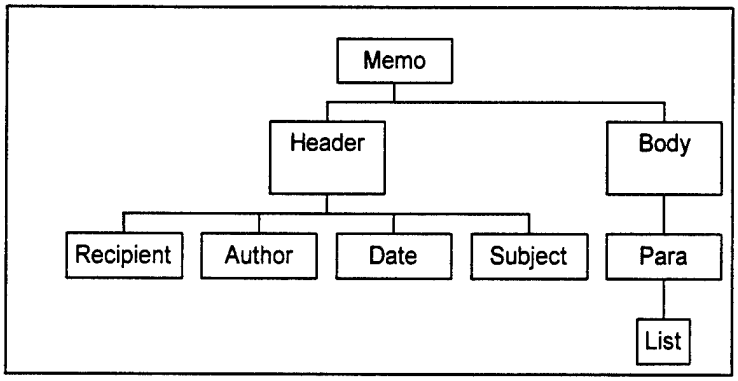

Figure 3: Structure of the Document Memo (Gilmore, 1993:213)

The DTD defines names of elements that are allowable, defines how often an element may appear, defines the order elements appear, defines tag attributes, and defines names of all entities that can be used. SGML parsers are programs that analyze and check that the markup in the document satisfies the rules defined by the DTD. Figure 4 provides a DTD for the memo pictured in Figure 2.

```
<!DOCTYPE memo(
<!ELEMENT memo          (header, body)>
<!ELEMENT header        (recipient+ , author+, date, subject?)>
<!ELEMENT recipient             (#PCDATA)>
<!ELEMENT author        (#PCDATA)>
<!ELEMENT date          (#PCDATA)>
<!ELEMENT subject       (#PCDATA)>
<!ELEMENT body          (para)+>
<!ELEMENT para          (#PCDATA 1 (list)+)+>
<!ELEMENT list          (item)+>
<!ELEMENT item          (#PCDATA)>
)>
```

Figure 4: Sample Document Type Definition for a Memo (Gilmore, 1993:214)

Content: Content is the information itself. Content includes titles, paragraphs, lists, and tables. The method for identifying the content within the DTD structure is called "tagging." Tags are inserted around the content in order to create an SGML document. These tags mark the beginning and end of each part of the structure. For example, a paragraph is preceded by a start tag of <para> and followed by the end tag of </para>. A tag contains a generic identifier that identifies the element type of an element (Heimburger, 1994: 243). Wright provided the following example of identifiers for the memo shown in Figure 2:

```
<memo>
<header>
<recipient> Students of SGML </recipient>
<author> Ms. Gilmore </author>
<date> 23 January 1996 </date>
<subject> Document Type Definition (DTD) for a Memo
</header>
<body>
<para>A good way to learn about SGML is to use it, beginning with simple
documents.</para>
<para>I am sending you a copy of the Memo DTD. Please note that the following
structural rules are built into the memo by the element definitions (called
declarations in SGML):</para>
<list> <item> There can be multiple recipients and authors.</item>
<item> There can be only one date</item>
<item> Providing a subject for the memo is optional</item>
<item> A body must follow a header.</item>
<item> A body must contain at least one paragraph (para)</item>
<item> Paragraphs can contain one or more list items but these are
       optional</item>
</list>
</para>
<para>I hope that you enjoy studying and using this DTD.</para>
</body>
</memo>
```

Figure 5: Memo Document Instance (Gilmore, 1993: 214)

The memo element in figure 5 contains a header, three paragraphs, a list, and several items. It is recommended when designing the DTD, the user first analyze the type of documents for which the DTD is required because a DTD is constructed so that it includes every element that is common to any given type of document (Heimburger 1994: 245; 1995, Stern 1995: 118). In order to ensure a concise DTD creation, representative related documents should be collected and analyzed to ensure their commonality in

structure and components. We should be able to easily identify the commonality with the structure and components of Air Force publications since AFI 37-160, Volume 1 governs how they are written and structured. For example, all publications must be in 10-point Times New Roman; have appropriate headings for each chapter, section, paragraph, attachments, and figures; and contain the proper italicized and bold fonts where needed.

An electronic publishing system must be able to handle both new and existing documents. Since SGML requires explicit and consistent structure, the DTD defined for the documents will impose rules that may not have been considered when the author created the original document (Gross, 1993: 220). Therefore, conversion to SGML can rarely be achieved one hundred percent automatically (Gilmore, 1993: 216; Maunder, 1994: 53; Gross, 1993: 220). Sometimes documents are available in printed form only or they exist electronically in varying formats in several different software applications. Older material can be retyped or converted with the use of scanners or optical character readers. The use of SGML allows documents to be acquired from many sources while the choice of what tools to use is left to the creator of the electronic source files. Much of the conversion can be done by software programs which can be purchased commercially off the shelf.

SGML Benefits

SGML, as described in ISO 8879, is an International Standard for document description published by the International Organization for Standardization and was released in 1986. SGML allows transfer of formatted text across hardware and software

platforms without the loss of content or format and allows reuse of information to include parts of old documents in new documents created for different purposes (Gilmore, 1993: 210; Alschuler, 1993: 208). This ability to share information is an improvement over the paper-based publication process because the information can be interchanged across heterogeneous systems regardless of the software application or hardware platform for which it was originally created (Turner, 1994: 53). A standardized approach to the electronic publishing system will allow us to transmit publications across varying machines and software applications without the need to re-key the data or filter it in order to use it again. This non-proprietary format allows the data to be used and reused in any way by any software application.

In an SGML document, both markup and content are represented with the ASCII character set (Gilmore, 1993: 211). (ASCII stands for American Standard Code for Information Interchange.) With paper documents, markup refers to the way an editor annotates comments and notes on a document to indicate how the document should be formatted. In electronic documents, markup refers to tags, a method for identifying the type of content within the document structure, which are interspersed with the text in the document. Tags are inserted around the content in order to create an SGML document. The markup tags separate a document's logical elements and may specify what functions should be performed (Wright, 1992: 280). For example, a markup tag for a list of items would indicate where the items begin, font to be printed in, and whether or not the items will be indented. Markup provides the procedural information and structure that is needed to manipulate and print documents (Wright, 1992: 280).

SGML provides a basis for identifying content and display factors for all types of media. Many experts in the field of SGML (Alschuler [1993], Davidson [1993], Gilmore [1993], Heimburger [1994], Lunemann [1995], Maunder [1994], Rath and Wiedling [1996], Stern [1995] and Wood [1995]) note that there are several benefits of using SGML. Benefits include:

1. **Portability of Information (Machine independence).** SGML allows for search, retrieval, manipulation, and transfer of data across various software applications, computer networks, and hardware platforms (open systems architecture). Since SGML is an international standard, documents can be exchanged with users who have different computer systems. "Which means your SGML document information will out live today's technology - and work with whatever comes along in the future (November 1995, WWW)."
2. **Expedited writing and production processes.** The structured approach to writing reminds the writer how to organize information. With the use of SGML, authors keep the document's content separate from style (October 1995, WWW). This allows authors to concentrate on content rather than appearance. End users can reuse pieces of the old document in a new document and save time by not having to recreate the information they need.
3. **Up-to-date information.** SGML allows for correction of errors or posting of updates to information quickly and easily. With a paper-based system, updates are time consuming and tedious, especially when copies of the publication are

located in several places. Failure to update all copies of the publication can result in operational errors and misinformation.

4. **Reusability**. SGML gives many applications access to the same set of information. Chunks of data can be identified with tags and used in many places as a task (October 1995, WWW). The task can appear in several formats; technical manuals, training guides, publications, etc. This chunking of data enables computers to manage and maintain the different use of the task in a single place.
5. **Information longevity**. There is no need to convert documents when software or hardware becomes obsolete (Gilmore, 1993: 211). SGML's standard file format allows information to always be available once the document is defined.
6. **Shareability**. Since SGML works with structured components of a document, entire documents are built out of information from various parts of the organization (October 1995, WWW). For a computer to be able to understand the structure of the document, it must be encoded through tags. Tags are really "place holders" which describe how the document is organized. An example of this is the Privacy Act statement that appears in many documents throughout the Air Force. A single tag in a document can pull in the Privacy Act statement and the statement can then be printed in any publication, eliminating duplication of information.
7. **Capability to perform searches and use of database technology**. SGML's well-defined structure makes it possible to perform complex searches in

documents (Lunemann, 1995: 345). In SGML, the sections of text are structured in such a way that text is marked with tags to help identify the structure. Database technology uses fields and records to identify information. SGML manipulates elements such as chapters, titles, headings and paragraphs instead of fields and records (Schettini and Alschuler, 1994: 73). Finding all publications in which the words "contractor compliance" are mentioned in the first paragraph is a complex task in traditional paper publications. However, SGML makes this a simple task.

These SGML benefits are intended to improve and expedite the development and use of information. The rationale behind the use of SGML is to make documents interchangeable so that creation, management, retrieval, and printing of information can be done using the hardware and software best suited for each task.

DoD Organizations using SGML

The Air Force has created the Automated Technical Order System (ATOS) to develop the publication and electronic delivery of Air Force maintenance manuals (Barron, 1989: 19). Since the maintenance manuals are produced at one of five Air Logistic Centers from electronic information compiled from several different contractors, ATOS allows the documents to be created within the requirements set by the military specifications regarding content and appearance. Another Department of Defense initiative is the Continuous Acquisition and Life-Cycle Support (CALS) program. CALS was developed to control the mounds of paperwork generated with the design,

development, manufacturing, and maintenance of weapons systems (Heimburger, 1994: 250; Barron, 1989: 21). The Department of Defense stores more than 200 million engineering drawings, and the Navy deals with 200,000 separate manuals associated with weapons systems (Barron, 1989: 21). The objective of CALS is to enable the integration of digital information for weapons systems acquisition, design, and support, to improve productivity and move toward electronic documentation (Brooks, 1994: 173). Defense contractors are now required to produce all documentation associated with weapons systems in SGML format to eliminate the mounds of paper associated with weapons systems.

The ability to display portions of a document by maintenance technicians under peacetime and wartime conditions was the driving force behind the Department of Defense's initiation of a project called Interactive Electronic Technical Manuals (IETM) (Reynolds, 1992: 266). The goal of IETM is to have an expert system determine the source of a problem, based on information provided by the maintenance technician, and display relevant portions of a document/technical manual on portable computers over a secure network. The information transmitted would allow the technician to repair the equipment without needing to refer to paper-based technical manuals. Currently, to perform flightline maintenance, technicians need to tote several technical manuals to the aircraft prior to performing any maintenance activity. If one technical manual used makes references to other manuals that are not readily available at the aircraft, the technician is required to obtain the additional technical manuals before continuing with

his maintenance activity. This laborious task of flightline maintenance can be simplified with the use of IETM.

Commercial Enterprises using SGML

The experience of commercial enterprises that are currently using SGML tell us that SGML offers many benefits for electronic publishing and provides flexibility and the power to create a single document that can be arranged in different formats to meet the end user's needs. Commercial enterprises have also found that SGML is cost- and time-effective. Historically, the prime adopters of SGML were defense contractors (Barron, 1989: 19; Heimburger, 1994: 250; Reynolds, 1992: 266). These contractors adopted SGML because of the CALS initiative to control the mounds of paperwork generated with the design, development, manufacturing, and maintenance of weapons systems. In the last few years many commercial organizations have recognized the benefits of SGML and have adopted it for their information management.

Ontario Hydro, Canada's largest utility, chose SGML to convert its 20,000 pages of documentation consisting of 11,000 pages of operator's manuals, 8,500 pages of training materials, and 300 pages of schematics to an electronic format (Rockley, 1993: 383). The on-line documentation is used by maintenance technicians to troubleshoot problems with their system and produce a printed copy of that portion of the technical manual associated with the problem. SGML allowed Ontario Hydro to create single-source documents that could be compiled into a number of different configurations to meet the needs of multiple users on multiple computer platforms (Rockley, 1993: 386).

Air Force aircraft mechanics and other specialty technicians would find the capability of displaying portions of electronic manuals an asset during troubleshooting.

Many organizations have adopted SGML to enable them to reduce their costs by publishing documentation electronically and printing documents on demand rather than stockpiling copies. Airlines and aircraft manufacturers are using SGML as the preferred tool to aid electronic delivery of aircraft documentation (Reynolds, 1992: 266). SGML technology is enabling development of electronic delivery of in-flight operations manuals the pilots view in the cockpit as well as maintenance manuals that the technicians can use on hand-held slates in maintenance hangars (Reynolds, 1992: 266).

Many organizations also want the capability of on-line access to provide support, technical information, and other electronic services (Glushko, 1993: 394). Silicon Graphics developed a system called IRIS InSight which contains several computer technical manuals. IRIS is an informational viewer with which the user can access and view on-line documentation provided by a computer vendor. SGML's standard for structured documentation allows IRIS to perform dynamic searches and link with graphics, audio, video and other digital media elements allowing the user to browse through the electronic manuals easily (Glushko, 1993: 396). Ontario Hydro also wanted the capability to provide on-line access to its technical manuals during troubleshooting. Since the on-line documentation for troubleshooting was structured using SGML, the technician can access all the electronic manuals at once and perform a search to find the necessary information. If the technician were required to find the information manually, he might need to have two or three manuals open at a time to perform the maintenance.

The Oxford University Press made a decision to computerize the 1933 edition of the Oxford English Dictionary (OED) in 1984 using SGML for the main purpose of developing a database for easy retrieval and searching of information (Fawcett, 1993: 380). The OED records the evolution of the English language since 1150 A.D. It also records the etymology of each word and provides many other types of information on the English language. The computerization of the OED allows for ongoing updates to over half a million words contained in the dictionary. Since end-users wanted to use the information on various equipment and in different ways, SGML provided the best solution for the conversion to electronic form. The electronic OED supports the capability to search for a text string in the dictionary in less than a second (Fawcett, 1993: 381). These queries would have been almost impossible to complete in hours or days with the paper-based OED. The electronic OED also allows users to extract portions of information for use in other documents using other software.

These examples suggest that SGML could help to eliminate the stockpile of publications and reduce the costs associated with printing documents across the Air Force. At any Air Force Publications Distribution Office there is a mountain of forms and publications that are stockpiled waiting for a user to request them. With the use of SGML, the documents can be printed on demand or made available electronically on CD-ROM, the Internet, or WWW.

SGML and World Wide Web Integration

The WWW allows organizations to reach out to large audiences with information that is changing rapidly and is time-critical (White Paper 4002-II, 1995: 6). It thus reduces the hurdles associated with hard copy distribution, such as high costs of reproduction and distribution and outdated or inaccurate information. The WWW gives end users instant access to documents and enables organizations to gain widespread visibility. Leading vendors in the SGML industry are exploiting their technology for Internet applications that will change the way organizations do business (White Paper 4002-II, 1995: 6).

SGML is playing an increasing role in information services available through the Internet. The WWW is a global hypertext system based on SGML. Documents available as readable files on the WWW are formatted in HyperText Markup Language (HTML), a subset of SGML. HTML is based on SGML, with its own Document Type Definition (DTD) and also relies on tags. The tags are read in the HTML file and the format is displayed on the user's computer system. HTML markup also provides text searches and retrieval capabilities through links to various, interrelated documents.

Since Air Force documents will be in the standard SGML format, they can be placed on the WWW for access. However, not all WWW browsers are capable of reading SGML documents. Since the documents on the WWW are formatted using HTML, SGML documents need to be converted to HTML with the use of a parser to meet the requirements of the specific browser. This conversion would actually reformat

the SGML document and transform it to fit the HTML DTD for use on the WWW. Another issue that must be addressed with respect to WWW publication is the size of the large SGML documents. Many WWW browsers seem to work best on documents which are a few pages in length (Freese, 1995). Freese (1995) recommends breaking the SGML documents into components and linking them. This chunking of information from the original document allows users to retrieve the components they desire without having to access a large document. Using the WWW capabilities will facilitate unilateral access to unrestricted standard publications and forms (and many other types of information) that Air Force personnel use on a daily basis. This new technology can provide the Air Force with information at its fingertips, but it also can have underlying problems.

"The United States is leading the world into a globally networked society, a true Information Age where information and economic value become nearly synonymous (Schwartau, 1994: 12)." While we help build a networked society, we should also be concerned and wary of the vulnerabilities associated with such new technologies. Schwartau says that information is intangible and does not have an immediately quantifiable monetary worth unless you lose it or someone alters it to suit his particular goals. For this reason, it is important that we take a closer look at the implications of information warfare as we migrate toward a paperless Air Force. While it is not the purpose of this thesis to research encryption schemes and risks associated with on-line access to information, future research into this area is warranted and discussed in Chapter Five.

Benchmarking

Benchmarking provides an opportunity to examine the best practices of other organizations and then implement changes based on the observations (Greengard, 1995: 64). An evaluation of benchmarking concepts and applications may provide useful information for determining if benchmarking can improve or streamline the Air Force's migration to an electronic publishing environment.

In its simplest form, benchmarking is the comparison of a given business function across companies. Benchmarking seeks to borrow and build upon proven techniques to improve efficiency, service, and output, enabling the organization to be the best in its class. Benchmarking allows for the sharing of information that can benefit everyone. It can streamline processes and help people do their jobs more efficiently. Samuel Greengard says that benchmarking is not about copying other companies best practices, it's about integrating the bits and pieces of useful information into a company and its culture (Greengard, 1995: 64).

Benchmarking was pioneered by Xerox Corporation in the late 1970's as a result of an effort to improve their copiers when Japanese-produced low-priced copiers entered the market. The most documented benchmarking project was Xerox's study of the L.L. Bean Company. Xerox saw the need to improve its warehousing and distribution process and turned to the industry leader L.L. Bean. Xerox visited their clothing mail order distribution center and discovered a warehousing and distribution process it could adapt to its copier business. Xerox benchmarked their distribution system and this allowed

Xerox to remain competitive in the copier market. Xerox has become a legend in benchmarking because of its pioneering efforts in the early 1980's (Harmon, 1992: 22).

Benchmarking Processes

Robert Camp, manager of benchmarking at Xerox, outlines the 10-step benchmarking process used by Xerox in his book, Benchmarking, The Search for Industry Best Practices that Lead to Superior Performance. Camp has become a prominent expert in benchmarking and his 10-step process is being used by major companies world wide (Harmon, 1992: 22). Camp organizes the 10-step process into five phases: Planning, Analysis, Integration, Action, and Maturity (see figure 6 below).

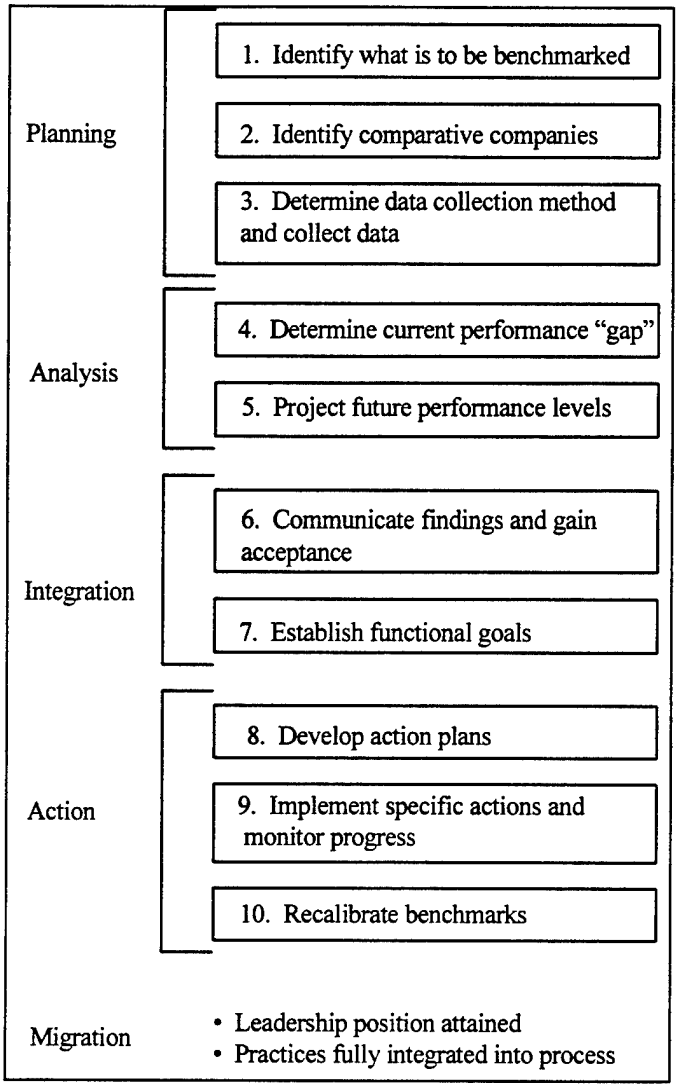

Figure 6: Benchmarking 10-Step Process (Camp, 1989: 17)

The objectives of the planning phase include determining what to benchmark, determining those people or operations to be used for comparison, and determining how the data will be collected. After planning the what, who, and how, data gathering and analysis must be accomplished. The second phase, analysis, includes a study of current business practices and an understanding of the practices for the benchmarking partners. This analysis will help determine the current performance gap. There are three types of

performance gaps: positive, negative, or parity. A positive gap exists when internal practices are superior. A negative gap exists when external practices are superior. Parity means that there is no significant difference. After the performance gap is identified, future performance levels can be projected during the next phase of benchmarking.

The third phase of benchmarking is called integration. During the integration phase, the results of the planning and analysis phases are communicated to personnel within the organization in order to gain operational and management acceptance. Upon completion of the integration phase, benchmark findings and operational goals are converted into action plans (phase four). These plans are implemented and progress monitored. When the industry's best practices are incorporated into all business practices, the last phase of benchmarking, maturity, will be reached (Camp, 1989: 19). If other organizations benchmark your own internal operations this is also confirmation that maturity has been reached.

This five phase, 10-step benchmarking process is being used by many companies to measure themselves against other organizations and solicit valuable input. Organizations that have utilized the 10-step benchmarking process include AT&T, Hewlett Packard, IBM, Ford, Goodyear, and more than 70% of the Fortune 500 companies (Port, 1992: 74; Greengard, 1995: 64; Harmon, 1992: 22). David Altany (1991) provides an excellent summary of the benchmarking process in his article. He says:

> Benchmarking's rapid ascent is surprising to some managers because the process itself is so straightforward and simple. A senior manager typically will start by deciding what part of the company to benchmark. The manager then instructs

specialists in that area to determine what company is the very best at that function and to start collecting data to exchange with that company. After analyzing the data, a strategic plan is developed to incorporate the most effective approaches used by the benchmarked company.

It is important to understand that the 10-step process can be applied to different types of benchmarking.

Types of Benchmarking

Camp (1989) identifies four types of benchmarking: internal, generic, competitive, and functional. Internal benchmarking involves benchmarking against an organization's internal operations and is often the first type of benchmarking used within an organization (Camp, 1989: 61; Harmon, 1992: 22). This type of benchmarking is also the simplest form and is used in large multi-division or international firms where there are similar functions in different operating units.

Generic benchmarking is benchmarking a process or function that is the similar, regardless of dissimilarities of industries (Camp, 1989: 65). This type of benchmarking has the potential to reveal the best-of-the-best practices. Generic benchmarking is the most difficult to gain acceptance and use but has the highest long-term payoff.

Competitive benchmarking is actually benchmarking against a direct product or service competitor. Camp states that although it is difficult to obtain information using the competitive benchmarking process, it should still be pursued (Camp, 1989: 63). The information obtained can be used by the other parties to understand the best practices that make their operations successful.

Functional benchmarking is the type of benchmarking used most often. This involves benchmarking against functional competitors or industry leaders, even if they are in dissimilar industries. According to Camp, functional benchmarking is most productive because organizations are interested in information on how the industry compares and what the best industry practices are. This natural interest in understanding practices allows for the organization to obtain information easier and share data.

Benchmarking Drawbacks

Although benchmarking is a powerful tool for organizational change, it is not as simple or straightforward as it may seem. Its complexity and implementation lead to potential problems. The first problem encountered by many benchmarking teams is determining the best-of-the-best companies and then getting those companies to share information. If a company lacks the commitment to find organizations that are truly the best a mediocre study will result. A partial solution to this drawback is the availability of the new International Benchmarking Clearinghouse (IBC). The IBC offers a wide range of benchmarking services, designed with information from companies that are recognized as industry leaders. The clearinghouse facilitates networking among members, collects and disseminates best practices through databases and case studies, and provides seminars and conferences related to benchmarking. However, there are two fees to join IBC - ranging from a $3,000 to $60,000 one time tax-deductible fee, and an annual fee ranging from $2,000 to $6,000 for basic service and networking capabilities (Harmon,

1992: 24; Port, 1992: 75). Management of each organization will need to decide if the information and services IBC provides is worth the price.

Companies should also be careful not to focus so closely on details that they completely miss the big picture (Lincoln, 1996: 33). Choosing the scope of the benchmarking study requires a thorough analysis of what is to be benchmarked, to whom or what will be compared, and how the data will be collected.

Last but not least, a lack of commitment from management can hinder the benchmarking process. Management must commit the necessary resources and time needed to conduct the benchmark study and allow the study to follow through to implementation. The benchmarking process may cover a period of month or years and commitment of management should be constant throughout the process. The organization must determine if the drawbacks of benchmarking can be overcome by its benefits.

Benchmarking Benefits

Benchmarking can benefit an organization in several ways (Camp, 1989: 66):

- It enables the best practices from any industry to be incorporated in the benchmarked function.
- It can provide motivation to the individuals required to implement and perform benchmark findings.
- Benchmarking breaks down the reluctancy to change operations. It has been found that people are more receptive to ideas that do not originate in their own organization.
- Benchmarking may identify a technological breakthrough that would not have been recognized for some time to come.
- Those involved with the benchmarking process find the professional contacts and interactions from benchmarking are invaluable and broaden their experience for future assignments.

Conclusion

The literature suggests that a standardized approach to an electronic publishing system will allow the Air Force to transmit publications across different platforms and software applications without the need to re-key the data or filter it in order to use it again. This approach should allow the Air Force to reduce the time it takes to publish Air Force documents, standardize document structure, eliminate storage of paper-based products, provide up-to-date information, and permit point-to-point availability of publications electronically on personal computers. Initial evidence suggests that SGML has proved to be a powerful data interchange standard that will meet the needs of the Air Force electronic publishing program to improve efficiency and effectiveness within the Air Force's publication process.

The investment in SGML makes sense for anyone who needs to publish several documents of the same type or the same document in several formats, to enforce document structure, or create documents that are continually updated or produced on demand. SGML also solves a problem most people do not know they have. Not until an organization reaches a crisis -- the inability to deal with backlog, the high cost of converting to a new system, or the failure to keep information accurate and up-to-date -- does the organization realize there is a problem (Heimburger, 1994: 258; Ensign, 1993: 393; Locke, 1992: 252). The cost of SGML depends on the value you place on information resources (Turner, 1994: 56). Many organizations view benefits as tangible or intangible; either will affect the view the organization has on the cost of using SGML. One organization may feel that the tangible benefit of saving money by decreasing costs

for printing and distribution outweighs the intangible benefit of having reliable, easily accessible information for employees. Since there is no cost/benefit analysis available on the conversion to SGML, we can examine commercial organizations through the benchmarking process to determine how they have dealt with the decision to implement SGML.

Benchmarking can be a frustrating and time-consuming task and it is important to develop a solid strategy and use methods that have successful records. This literature review on benchmarking suggests that the Air Force can profitably use the experience of commercial enterprises to help during its migration to an electronic publishing environment. Benchmarking will provide a useful tool in examining the practices that have been employed so that the Air Force can integrate pieces of useful information into its plan.

With the proper use of research, the Air Force can exploit SGML's benefits fully in its migration from paper to electronic publishing. We need, therefore, to explore in more depth the commercial enterprises who have elected to use SGML and evaluate the findings to determine if the benefits of SGML can be implemented by the Air Force. The Air Force can definitely gain much needed information about SGML's benefits and capabilities from commercial enterprises through the benchmarking process. The information gathered will assist the Air Force in incorporating the best practices into its operations and capitalize on the benchmarking benefits during their migration to SGML.

III. Methodology

Introduction

This research effort is an exploratory, qualitative study which defines the components and benefits of SGML and the benchmarking application and process. Research in this area for the Air Force does not currently exist. Therefore, qualitative research is the best method for this study. Because this study is also exploratory, it fits Cooper's definition - "Exploratory studies tend toward loose structures with the objective of discovering future research tasks. The immediate purpose of exploration is usually to develop hypotheses or questions for further research" (Cooper, 1995: 115). In order to gather the necessary information and data on SGML and benchmarking for this study, the qualitative techniques used include secondary data sources (literature review) and primary data sources (questionnaire).

Secondary Data Collection

The first step in this study was an initial, exploratory review of existing literature to determine the major areas of interest in SGML and benchmarking to be considered for further research. This initial review provided major sources for information gathering; the components and benefits of SGML and basic benchmarking concepts. The amount of information available during this initial review was overwhelming. Over 2,000 sources exist for SGML-related information and more than 600 sources for benchmarking. These sources were available by searching AFIT's on-line catalog, the FirstSearch on-line reference system (an electronic database with over 28 million records), and the WWW.

The WWW provided an excellent address, URL: http://www.sil.org/sgml/biblio/html, to an SGML bibliography document that contains references and abstracts for over 1000 books and articles. This bibliography was used as the major source for the collection of secondary data pertaining to SGML for this study.

The review of periodicals, books, government documents, and the WWW developed valuable information on SGML components and benefits and benchmarking techniques. The enormous amount of information available on SGML and benchmarking, made it necessary to pare down the literature and search only for information that was pertinent to the study. From the information gathered during the initial review, a checklist was developed to ensure that the sources obtained and information collected was deemed necessary for the research. The development of a checklist provided a systematic process to review the literature. The checklist was divided into two parts. One section for SGML and the other for benchmarking. The checklist for SGML included the following areas for research: components of SGML, organizations using SGML, uses of SGML, conversion/authoring methods, benefits of SGML, and training issues. The benchmarking checklist was composed of the following: definition of benchmarking, purpose of benchmarking, the benchmarking process, applications of benchmarking, uses of benchmarking, list of organizations that are the best in the industry, and benefits of benchmarking.

Upon completion of the initial exploratory review, the secondary data collection consisted of two separate literature reviews. The first literature review was conducted to gather information on SGML. This literature review established criteria that can be used

by Air Force publishing centers to ensure their migration to electronic publishing exploits the benefits of SGML.

The second literature review revealed the application and process of benchmarking. An extensive search of existing literature was performed to identify benchmarking principles. Although benchmarking requires using the best industry practices, no such ranking of electronic publishing experts using SGML existed in the literature.

Primary Data Collection

Building on the background provided by the literature review, a questionnaire was designed based on this information to develop Air Force-specific information. According to Cooper, "While published data are a valuable resource, seldom is more than a fraction of the existing knowledge in a field put into writing" (Cooper, 1995: 119). It is for this reason that a questionnaire was developed based on the literature review. The questionnaire was used to seek information from organizations that have been using SGML for their electronic publications. The information gathered from these questionnaires was used to confirm and elaborate the findings of the secondary data and formed the framework for development (addressed in Chapter Four) of the findings.

Population (Sample)

The population of interest for the study comprises commercial enterprises who have converted documents to SGML or have been using SGML within their organization

to author documents. A sample of the population was developed by reviewing existing literature, the WWW, and the SGML Newswire Hit List for any commercial enterprises with that association to SGML. The SGML Hit List is a collection of case studies on SGML compiled by Avalanche Development Company which includes a reference list of several industry analysts, industry publishers, and organizations that are using SGML. The organizational contacts in the reference list are SGML users who have volunteered to discuss projects they are undertaking which use SGML conversion and/or authoring techniques. The WWW search was conducted using an Internet search engine to elicit web pages that contained the words "SGML" and "Users," "SGML" and "Customers," or "SGML" and "Implementation." This search returned over 600 web addresses. However, not all the web addresses referenced web pages that contained useful information on commercial enterprises. To validate a useful web page, a cursory review of the web site was performed to ensure the organization was an enterprise that had converted documents to SGML or had been using SGML to author documents. Only those web pages which contained a mailing address or fax number for the commercial enterprise were kept and the enterprise was then considered for the questionnaire sample. As Patton points out, determining the appropriate sample size for a qualitative study is very difficult. "It depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources" (Patton, 1990: 184).

The sample size for this study is 65 commercial enterprises (see Appendix B for list of enterprises contacted). Sample size was not deemed as important as industry-wide

representation by organizations. An important aspect of this study was to insure that enterprises from all business sectors were considered. This cross-sectional representation of enterprises was obtained by searching the WWW and literature. The enterprises were selected based on whether or not they have used SGML to convert or author documents within their organization. Although individual interviewee responses are reported for each question, these responses are considered to be representative of the respective organization. This representation was deemed appropriate since the most experienced or knowledgeable person involved with the organization's migration and implementation of SGML provided a reply to the questionnaire.

Data Collection Method

The time and distance constraints in collecting the necessary information from the commercial enterprises limited the data collection method to the use of a questionnaire. The commercial enterprises identified to be sampled are located across the country. For this reason, a questionnaire was mailed or faxed to the selected commercial enterprises in May 1996 and each organization was asked to reply within 30 days of receiving the questionnaire. The questionnaire gathered data on demographics of the organization, the organization's use of SGML, products/tools used for conversion or authoring of SGML documents, training associated with SGML, and the problems and benefits of SGML's implementation. The complete questionnaire is reproduced in Appendix C.

The questionnaire was reviewed prior to conducting the study. A member of the Information Architecture Branch, Air Force Materiel Command Headquarters, was asked

to evaluate the questionnaire prior to it being sent out. Suggestions and comments were used to revise and improve the questionnaire.

Design of Instrument

The questionnaire sent to the commercial enterprises was an exploratory instrument designed to provide information about the organization and its implementation and use of SGML. The questionnaire did not aim to establish causal relationships but was intended to collect information on SGML. The questionnaire also collected information about the organization to determine if the organization could be considered as one which the Air Force could benchmark.

The questionnaire consisted of structured and unstructured questions. The first five questions were designed to elicit demographics about the commercial enterprise. The next 17 questions were quick response questions designed to merely provide background information about the enterprise's use and capabilities of SGML. These questions were related to number of pages converted to SGML, number of documents authored, and purpose for using SGML. Other questions were designed to provide information on specific topics related to advantages, training, and benefits associated with using SGML. Question 23 was open-ended, unstructured, and designed to elicit any important information regarding the organizations' experiences using SGML.

Analysis

As stated earlier in this chapter, an exploratory study is used to shed light on areas previously not understood. Based upon the data collected, answers to the individual

research questions are reported in Chapter Four. These answers are then specifically related to the research questions in order to facilitate analysis of the overall implications of the Air Force's conversion of publications to SGML. Finally, the findings of the analysis made from individual research question analysis will be integrated in Chapter Five.

Summary

The methodology used in this study is mainly qualitative in nature. Secondary data collection was used to determine the concepts and benefits of SGML. The information gathered from the literature review should be useful in determining if Air Force publishing centers can use the full potential of SGML in their migration to electronic publishing. Further, a literature review of benchmarking was accomplished to establish the concepts and applications of benchmarking. This information will be useful in determining if the Air Force can benchmark a commercial enterprise to improve or streamline their migration to an electronic publishing environment.

Primary data analysis was used to confirm and elaborate the findings of the secondary data analysis. The questionnaire was administered to seek information from commercial enterprises accomplished in the field of SGML and electronic publishing. The information gathered from these questionnaires was integrated with the secondary data identified in the literature review. This chapter describes the methodology used by the researcher to help answer the investigative questions from Chapter One. The next chapter will present the results of the literature review and questionnaire.

IV. Research Findings

Introduction

This chapter presents the research findings from the literature review and the questionnaire responses. The findings from the literature are divided into two sections. The first section covers the data collected on SGML and the second addresses benchmarking concepts. The analysis of the questionnaire results begins with an analysis of the responses to the demographic questions (questions 1-5). Next, a comparison of the responses concerning the commercial enterprises' use, capabilities, and benefits of SGML is presented (questions 6-22). Finally, the open-ended responses to question 23 are analyzed to determine if the respondents provided additional information about their organization's benefits and experiences using SGML that would be of assistance to the Air Force publication centers during their migration to SGML.

SGML Literature Review Findings

The Air Force Electronic Publishing Program, which began in 1993, has proposed the migration of all Air Force paper products in the publishing inventory to a standard electronic product. Since the time of this program, the Air Force has converted many of its publications from paper-based publishing to a more advanced, search capable CD-ROM format which provides on-line access to publications. The Electronic Publishing Program has proposed the integration of SGML into the publication process. However, the Air Force does not have the necessary research in the area of SGML to ensure the migration fully exploits all of SGML's capabilities and benefits. The literature review

presented in Chapter Two discusses the potential of SGML, the benefits of SGML, and reviews commercial enterprises that have or are currently using SGML in their electronic publishing environment.

Benefits of Using SGML

The major sources for the literature review included journals, books, and the WWW. Seventeen of the 35 sources referenced, addressed the benefits of SGML. The seven most common benefits include machine independence, capability to perform searches or the use of database technology, reuseability, expediting the production/writing process, information longevity, shareability, single source input for many different output capabilities, and up-to-date information. The following table presents the percentage of sources that discussed each benefit.

Table 1: Percentage of SGML Benefits Cited in Articles

Benefits of using SGML	**Percentage of articles citing benefits**
Provides machine independence	43%
Permits search and/or database technology	40%
Allows reuseability of information	26%
Expedites production/writing process	20%
Increases the longevity of information	14%
Permits shareability of document components	11%
Allows for different output capabilities from one single source document	11%
Provides up-to-date information	6%
Improves data quality	6%
Allows for security of information	3%
Reduces operating costs/failures	3%
Improves data integrity	3%
Allows merging layouts of several authors	3%
Improves market gains	3%

More than 50 percent of the sources emphasized the importance of content being separate from style. The components of SGML are important to understand prior to the Air Force undertaking the migration. Thirty-seven percent of the sources collected addressed the components of SGML. Some authors went into great detail while others just covered only the most basic characteristics. Seventeen percent of the literature sources referenced Dr. Goldfarb's book, The SGML Handbook, as a source of valuable information, one that is essential to implementing SGML. This book provides the full text of ISO 8879, a detailed overview of SGML, and tutorial and reference material.

Commercial Enterprises using SGML

In evaluating whether SGML can have a beneficial impact on the Air Force's ability to migrate from paper to electronic publishing, a review of commercial enterprises was conducted. According to Barron, Heimburger, and Reynolds, the prime adopters of SGML were defense contractors (Barron, 1989: 19; Heimburger, 1994: 250; Reynolds, 1992: 266). However, 8 of the 35 sources collected provided information on commercial enterprises other than defense contractors. A WWW search returned over 600 web sites that contained information related to SGML and users or implementers. Table 2 provides a list of several organizations and their use of SGML.

Table 2: Commercial Enterprises and Their Use of SGML

Commercial Enterprise	Use of SGML
Airline Industry	Maintenance documentation/in-flight operating manuals
American Assoc. of Publishers	Electronic manuscript preparation
Automotive	Service/repair information
Computer hardware and software	On-line help/user documentation
Defense organizations	Maintenance contract/on-line technical manuals
Electronics	User documentation
European Aerospace	Documentation
Institute of Electrical Engineers	Producing alternative forms of publications from the same source
Internet	World Wide Web
Ontario Hydro Utility Plant	On-line technical manuals
Oxford English Dictionary	Search/database retrieval
Shell UK Exploration & Production	Management of safety manuals
Silicon Graphics - IRIS InSight	On-line documentation
Telecommunication Industry Forum	Reuse of technical information

Benchmarking Literature Review Findings

The literature reviewed the concepts and applications of benchmarking. The major sources for information included journals and books. Additionally, the WWW provided many sources of companies that are available to provide services to organizations that wish to undertake benchmarking.

All sources indicated that there is a standard approach to benchmarking. However, no single model has emerged to guide prospective benchmarkers. The several existing models include anywhere from four to ten steps. Fifty percent of the sources collected for the benchmarking literature review referenced Xerox's 10-step benchmarking process. Robert Camp's book, Benchmarking, The Search for Industry Best Practices that Lead to Superior Performance, best identifies and outlines the Xerox process.

The literature suggests that all benchmarking projects have the same objectives: determine what to benchmark, determine the people or operations to be used for comparison, and determine how the data will be collected. Forty percent of the journal articles indicated that many companies that perform benchmarking discover that direct competitors are not always likely to know the best way to perform a function or process. Rather than comparing to direct competitors in the same industry, organizations can examine practices and methods of organizations in dissimilar industries. Comparing organizations in dissimilar industries allows the benchmarking organization to find industry-wide best practices *not* the best practices within a particular business sector.

Questionnaire Findings

The questionnaire was sent to 65 commercial enterprises to obtain information about the organization and its implementation and use of SGML. Twenty SGML users provided responses to the questionnaire (a list of enterprises are provided in Appendix D). This corresponds to a response rate of 31 percent. In addition to the twenty responses, three questionnaires were returned but their responses were not used because the organizations were not users of SGML; rather they provide consulting and conversion services to organizations that are interested in SGML.

The questionnaire consisted of 22 quick response type questions in which the responder was asked to check a box or enter a reply in a place available for "other" responses. The last question was open-ended and unstructured. In the following discussion, the responses to the questionnaire are preceded by the percentage which

corresponds to the organizations' responses. These figures do not indicate a level of importance for any one reply over another; rather they indicate the level of frequency in which organizations replied to the question. The consolidated results of the questionnaire are provided in Appendix E.

Demographics

There were five questions in the questionnaire used to gather background information on the organizations providing responses. The demographic questions, along with their consolidated responses, were:

1. What are the primary activities of your organization?

Table 3: Primary Activity of Responding Organizations

Primary of Activity of the Organization	Percentage of Respondents that Selected Each Activity
Services	35%
Business/Industry	30%
Publishing	25%
Computing	20%
Other (reference Appendix E for respondents' comments)	20%
Manufacturing	15%
Education	15%
Government	10%
Communications	5%
Retail	5%
Research	0%
Public Administration	0%
Wholesale	0%

The percentage of responses do not add up to 100 percent because the respondents could select more than one type of activity for their organization. The responses indicate

there is a wide variety of industries represented with Services and Business being the two most prominent.

2. How many employees are in your organization?

Table 4: Number of Employees in the Organization

Number of Employees in the Organization	Percentage of Respondents in Each Category
More than 1000 employees	50%
26 to 250 employees	30%
1 to 25 employees	10%
251 to 500 employees	5%
501 to 1000 employees	5%

Responses from the commercial enterprises indicate that 50 percent of the organizations employ more than 1000 people.

3. How long has your organization been using SGML?

Table 5: Number of Years the Organization has been using SGML

Number of Years the Organization has been using SGML	Percentage of Respondents in Each Category
3 to 5 years	30%
5 to 7 years	30%
2 to 3 years	20%
More than 7 years	10%
Less than 1 year	5%
1 to 2 years	5%

The responses indicate that the modal number of years for using SGML range between 3 to 7 years. Thirty percent of the organizations replied by stating they have been using SGML 3 to 5 years and 30 percent stated they have been using SGML for 5 to 7 years.

4. How many people in your organization are using SGML?

Table 6: Number of Employees using SGML

Number of Employees in the Organization that are using SGML	Percentage of Respondents in Each Category
25 employees	25%
5 employees	15%
20 employees	10%
50 employees	10%
200 employees	10%
10 employees	5%
12 employees	5%
15 employees	5%
40 employees	5%
100 employees	5%
1750 employees	5%

Replies for number of employees in the organization using SGML range from 5 to 1750. The mean number of employees using SGML is 131.

5. How many outsourcing contractors in your organization are using SGML?

Table 7: Number of Contractors using SGML

Number of Contractors in the Organization that are using SGML	Percentage of Respondents in Each Category
No contractors	50%
1 contractor	15%
2 contractors	10%
4 contractors	5%
6 contractors	5%
8 contractors	5%
10 contractors	5%
15 contractors	5%

Replies for number of contractors in the organization using SGML range from 0 to 15. The mean number of contractors using SGML is 2.5.

Use of SGML

Thirteen questions in the survey were constructed to define why the organization is using SGML, the benefits they have realized, and problems they have encountered. Each question will be addressed individually starting with question 6.

6. Why have you chosen to use SGML?

Table 8: Reasons for Using SGML

Reasons for Using SGML	Percentage of Respondents Listing Each Category
Required to meet demand of customers	60%
Change in information technology	50%
Compliance to a standard or contractual agreement	30%
Other (reference Appendix E for respondents' comments)	30%
Competition	10%
Required by upper management	5%

The percentage of responses do not add up to 100 percent because the respondents could select more than one type of activity for their organization. The responses indicate that customer demand and a change in information technology are the main drivers for the conversion to or implementation of SGML.

The next two questions were related to the actual number of pages each organization has converted to SGML and number of documents authored per year using SGML. The questions were:

7. How many pages have you converted to SGML?

Table 9: Number of Pages Converted to SGML

Number of Pages Converted to SGML	Percentage of Respondents in Each Category
More than 2000 pages	90%
101 to 500 pages	5%
1001 to 2001 pages	5%
None	0%
1 to 100 pages	0%
501 to 1000 pages	0%

The responses indicate that 90 percent of the respondents have converted more than 2000 pages to SGML. The results of this question suggest that the "percentage of pages converted" would be have a better indicator of how many pages these commercial enterprises have converted. At the time the questionnaire was developed, the literature indicated that organizations have converted approximately 500 to 1000 pages. The reason for the high percentage of organizations converting more than 2000 pages could be due to the experience gained over the past couple of years and the availability of software applications for converting legacy documents to SGML.

8. How many SGML documents do you author per year?

Table 10: Number of Documents Authored per Year

Number of Documents Authored per Year	Percentage of Respondents in Each Category
25 or more documents	65%
11 to 24 documents	20%
None	5%
1 to 10 documents	5%

The percentage of responses do not add up to 100 percent because one respondent did not reply. The responses indicate that 65 percent of the respondents author more than 25 documents per year.

The next question seeks to determine the reason the commercial enterprises are using SGML. The question is as follows:

9. For which of the following do you use SGML?

Table 11: Uses of SGML

Uses of SGML	Percentage of Respondents Listing Each Item
Electronic books	90%
On-line distribution	75%
HTML/WWW	65%
CD-ROM	65%
Intent for World Wide Web use	60%
Ensuring structural consistency	60%
Legacy documents	55%
Information preservation	55%
Document management	35%
Varied platform use	35%
Storage	25%
Other (reference Appendix E for respondents' comments)	15%

The percentage of responses do not add up to 100 percent because the respondents could select more than one use of SGML. The top two uses of SGML are Electronic Books and On-Line Distribution. However, the distribution of responses indicate there is not one single use for SGML, but there are many uses and reasons for an organization to decide to implement SGML. These commercial enterprises may have converted to SGML for one initial purpose and after the conversion other uses come to the forefront.

The next two questions are directly related to the type of documents and information contained within the documents of the commercial enterprise. Question 10 was:

10. For what types of documents are you using SGML?

Table 12: Types of Documents for which SGML is Used

Type of Documents for which SGML is Used	Percentage of Respondents Listing Each Item
Technical Manuals	90%
User Manuals	65%
Electronic Data Books	40%
On-line Electronic Database	35%
Books/Publishing	25%
Regulations/Publications	25%
Other (reference Appendix E for respondents' comments)	15%
Letters/Memos	10%
Request for Proposals (RFP)	5%

The percentage of responses do not add up to 100 percent because the respondents could select more than one type of document for which they are using SGML. The responses indicate that the main types of documents for which SGML is used are technical manuals and user manuals. The typical letter and memo is not the product driving the use of SGML; only 10 percent of the respondents are using SGML for these types of documents. The variety of styles and the number of authors creating letters and memos may be a couple of reasons SGML isn't as popular.

11. What types of information are contained in your documents?

Table 13: Types of Information Contained in Documents

Type of Information Contained in Documents	Percentage of Respondents Listing Each Category
Standard Text	100%
Figures and Graphics	100%
Tabular Information	95%
Links and Cross References	80%
Math and Equations	60%
Other (reference Appendix E for respondents' comments)	5%

The percentage of responses do no add up to 100 percent because the respondents could select more than one type of information that is contained in their documents. The distribution of responses across all types of information indicate that the commercial enterprises are taking full advantage of SGML and using it for more than just standard text. In fact, virtually all types of information are being incorporated into the documents by these commercial enterprises.

Question 12 was asked in order to find out how many commercial enterprises are actually converting legacy documents (existing paper-based products) to SGML or if they are simply converting documents that already exist in some electronic format.

12. Are your source documents hardcopy (paper) or electronic?"

Table 14: Type of Source Documents

Type of Source Documents	Percentage of Respondents Listing Each Category
Electronic	65%
Hardcopy (paper)	60%

The percentage of responses do not add up to 100 percent because the respondents could select more than one type of source document. The responses indicate that the organizations are in fact converting legacy documents and electronic documents to SGML. This information is very useful for the Air Force because the majority of their publications are legacy documents or electronic documents created using Word or WordPerfect software.

The next question was developed in order to identify what type of authoring is performed by the commercial enterprises. Question 13 was:

13. What percent of authoring is done with the following methods?

Table 15: Authoring Methods

Authoring Methods	Percentage of Respondents Listing Each Category
Direct Authoring	70%
Scanning from a printed source using OCR	48%
Translating from an unstructured electronic source	34%
Generating from a database or spreadsheet	18%
Keyboarding from a manuscript	15%

The percentage of responses do not add up to 100 percent because the respondents were asked to provide the percentage of authoring methods they employ and an mean was computed based on the replies received. The responses indicate that the majority, 70 percent, of new documents are created from scratch using SGML and 48 percent are scanned from an existing source.

Question 14 was asked to determine what environment the commercial enterprises are implementing and using SGML. The next question was:

14. What operating systems do you use?

Table 16: Types of Operating Systems

Types of Operating Systems	Percentage of Respondents Listing Each Category
MS-Windows	80%
UNIX	70%
MS-DOS	55%
Macintosh	25%
Other (reference Appendix E for respondents' comments)	15%

The percentage of responses do not add up to 100 percent because the respondents could select more than one type of operating system they use. The responses indicate that 80% of commercial enterprises are using MS-Windows. This is the same environment that the Air Force publications centers operate in.

The next question was developed in order to compile a list of products/tools that could be considered for the Air Force's migration to SGML. Question 15 was:

15. Which of the following products/tools do you using for conversion/authoring?

Table 17: Product/Tools used for Conversion and Authoring

Products/Tools used for Conversion and Authoring	Percentage of Respondents Listing Each Category
Other (reference Appendix E for respondents' comments)	60%
Author/Editor	50%
DynaText	45%
In-House Applications (reference Appendix E for respondents' comments)	45%
FastTag	40%
SGML Hammer	20%
InContext	15%
Interleaf	15%
FrameBuilder	10%
Intellitag	5%
Write-It	5%
MS Author	5%
HyMinder, SGML Tagger, SGML/Search	0%
WorldView, Application Builder	0%
IBM SGML Translator	0%

These percentage of responses do not add up to 100 percent because the respondents could select more than one type of product/tool they use for conversion and authoring. The distribution of responses indicate there are several different products and tools available for conversion and authoring.

Question 16 was asked in order to confirm and elaborate the information collected during the literature review on the benefits of using SGML. Question 16 was:

16. What do you believe are the benefits of using SGML?

Table 18: Benefits of Using SGML

Benefits of Using SGML	Percentage of Respondents Citing Each Benefit
Permits portability of information	95%
Allows reuseability of information	85%
Increases productivity	60%
Permits shareability of document components	60%
Permits flexibility	55%
Improves data integrity	55%
Increases the longevity of information	50%
Provides better data control	45%
Permits cost efficiency	40%
Provides up-to-date information	35%
Improves participation in global markets	35%
Other (reference Appendix E for respondents' comments)	15%
Allows for security of information	5%

The percentage of responses do not add up to 100 percent because the respondents could select more than one benefit of using SGML. The responses indicate the two most prominent benefits of using SGML are portability of information and reuseability. However, the distribution of responses indicate that organizations are gaining many benefits from using SGML.

The next question was asked in order to gain insights into whether the commercial enterprises have conducted a cost/benefit analysis. Question 17 was:

17. Has your organization conducted a cost/benefit analysis of SGML in general?

Table 19: Cost/Benefit Analysis

Conducted a Cost/Benefit Analysis?	Percentage of Respondents in Each Category
Yes	25%
No	70%

The percentage of responses do not add up to 100 percent because one respondent did not reply. Twenty-five percent of the commercial enterprises have conducted a cost/benefit analysis. The comments provided about their results do not give specific dollar figures for the cost savings (reference Appendix E for respondents comments). The responses indicate intangible benefits such as data reuse and accessibility to information in multiple output formats.

Question 18 was asked to determine any problems associated with the implementation and use of SGML. The question asked:

18. Which of the following have been problems you've encountered using SGML?

Table 20: Problems Encountered Using SGML

Problems Encountered Using SGML	Percentage of Respondents Listing Each Item
Training	40%
Conversion	40%
Tagging of documents	25%
Authoring	20%
Other (reference Appendix E for respondents' comments)	20%

The percentage of responses do not add up to 100 percent because the respondents could answer with more than one problem they have encountered when using SGML.

The responses from the commercial enterprises indicate that problems exist in several areas of SGML application. The top two problems are in the area of conversion and training. Converting documents from paper has been identified as a major slow down in the process of using SGML. The main reason as indicated by the respondents' comments is that the source (paper-based product) is unstructured and requires "manual cleaning up" prior to SGML tagging.

SGML Training

Questions 19 and 20 asked about the organizations' training efforts and training plan for SGML. The two questions were:

19. Does your organization provide SGML training to its employees?

Table 21: Is SGML Training Provided to Employees

Is SGML Training Provided to Employees?	Percentage of Respondents in Each Category
Yes	60%
No	35%

If yes , do you feel this training has enhanced employees' use of SGML?

Table 21.1: Does Training Enhance Employees' use of SGML

Does Training Enhance Employees' use of SGML?	Percentage of Respondents in Each Category
Yes	100%
No	0%

If no, do you believe training would be beneficial?

Table 21.2: Would Training be Beneficial?

Would Training be Beneficial?	Percentage of Respondents in Each Category
Yes	85%
No	0%

The percentage of responses to question 19 do not add up to 100 percent because one respondent did not answer this question. The responses indicate that sixty percent of the commercial enterprises provide training to their employees and all feel that it has enhanced their employees' use of SGML. Thirty-five percent do not provide training and of these 35 percent, 85 percent believe training would be beneficial to enhancing their employees' use of SGML.

20. Does your organization have a training plan for conversion/authoring SGML documents?
(For example: Basic SGML training for new employees, in-depth training after 3 months on the job, continuing training after 6 months on the job)

Table 22: Does your Organization have a Training Plan?

Does your Organization have a Training Plan?	Percentage of Respondents in Each Category
No	50%
Yes	45%

If yes, how effective is this plan?

Table 22.1: Is Training Plan Effective?

Is Training Plan Effective?	Percentage of Respondents in Each Category
Meets needs of organization	55%
Very Effective	44%
Not Effective	0%

If no, would a training plan be beneficial for your organization?

Table 22.2: Would a Training Plan be Beneficial?

Would a Training Plan be Beneficial for your Organization?	Percentage of Respondents in Each Category
Yes	80%
No	10%
Not Sure	10%

The percentage of responses to question 20 do not add up to 100 percent because one respondent did not answer this question. Forty-five percent of the respondents have a training plan for the conversion/authoring of SGML documents and 44 percent feel the training plan is "very effective," while 55 percent said it "meets the needs of the organization." Fifty percent of the commercial enterprises do not have a training plan and of these enterprises, 80 percent believe one would be beneficial. Ten percent of the organizations that do not have a training plan said it would not be beneficial and one respondent was not sure if a training plan would help.

SGML Resources

Question 21 was asked in order to gain more information about the resources that the commercial enterprises have used to learn about SGML. Question 21 was:

21. What resources have you used to learn about SGML?

Table 23: Resource for SGML Information

Resource Available to Learn About SGML	Percentage of Respondents Listing Each Item
Books	80%
Practice/Experience	80%
Training	70%
Co-Workers	65%
User Groups	50%
Newsletters	40%
Internet	40%
Other (reference Appendix E for respondents' comments)	30%

The percentage of responses do not add up to 100 percent because the respondents could answer with more than one resource they use to learn about SGML. The responses indicate that most organizations have learned about SGML through books and practice/experience. However, the distribution of responses shows that there are a wide variety of sources available.

Question 22 was asked to gain permission to contact the commercial organizations in case additional information was needed or clarification of responses was necessary. The question was:

22. I may wish to contact you for further information regarding this questionnaire or SGML. May I call you?

Table 24: Permission to Contact for Additional Information

May We Contact you for Additional Information?	Percentage of Respondents in Each Category
Yes	100%
No	0%

All of the commercial enterprise responded "yes" indicating they would provide additional information if necessary.

Additional Information from the Enterprises

The last question was open-ended, unstructured, and designed to elicit any comments the commercial enterprises would like to add. Question 23 was:

23. Is there any additional information you would like to provide?

These additional comments (reference Appendix E) provided useful information on the organizations' use of SGML and problems they have encountered. These comments provide additional information in answering research questions 3 and 4: "Will those commercial enterprises using SGML be able and willing to provide information regarding the benefits of using SGML? If so, what are the benefits of using SGML in the commercial enterprise?" and "What underlying problems or concerns are associated with these commercial enterprises achieving the benefits of SGML?"

Summary

The analysis of the literature review and the responses to the questionnaire provide valuable information and insights concerning SGML and benchmarking. Despite the small sample size, the results of the questionnaire confirmed and elaborated the findings of the literature review. In addition, it appears that benchmarking can be used by the Air Force to improve or streamline their migration to an electronic publishing environment. Chapter Five discusses the findings from this chapter and provides recommendations for future research.

V. Discussion, Conclusions, and Recommendations

Introduction

This study addressed six research questions used to determine the capabilities and benefits of SGML that the Air Force can use to enhance future publications and to conclude if benchmarking can improve or streamline the Air Force's migration to an electronic publishing environment. A discussion and conclusions drawn from the five research questions presented in Chapter One will be presented first, followed by recommendations. Finally, areas for future research will be presented.

Research Question One

Research question one asked, "What are the benefits of using SGML that will be of value to the Air Force?" In order to understand the importance of SGML and its value to the Air Force electronic publication environment, it was necessary to identify the benefits of SGML from existing sources. The literature review identified the seven most common benefits to an organization using SGML as the capability to prepare documents that (1) permit portability of information, (2) allow end-user to perform searches, (3) permit reuseability of information, (4) expedite the production/writing process, (5) provide information longevity, (6) permit shareability of information, and (7) provide for various output capabilities from a single-source input. A discussion of each benefit is provided in Chapter Two along with a table (Table 1) presenting the percentage of sources that discussed each benefit in Chapter Four.

Research Question Two

Research question two asked, "Which commercial enterprises are using SGML?" The information gathered from the literature review and the 20 questionnaires indicates a wide representation of industry organizations that have converted to or are currently using SGML within their organization (reference Table 2 and Table 3). Historically, the prime adopters of SGML were defense contractors (Barron, 1989: 19; Heimburger, 1994: 250; Reynolds, 1992: 266). However, the literature suggests that organizations from other industries are adopting SGML for their electronic publications. To determine what types of organizations are actually implementing SGML, the questionnaire was sent to defense contractors and service- and business-related industries. All these industries are represented in the responses presented in Chapter Four.

Research Question Three

The following discussion addresses research question three. The question was: "Will those commercial enterprises identified in research question two be able and willing to provide information regarding the benefits of using SGML? If so, what are the benefits of using SGML in the commercial enterprise?" Table 18 in Chapter Four presented the responses from the commercial enterprises on the benefits they have achieved using SGML. The following figure provides an analysis of the percentage of respondents citing benefits of SGML (Table 18) to the percentage of articles citing the benefits of SGML (Table 1).

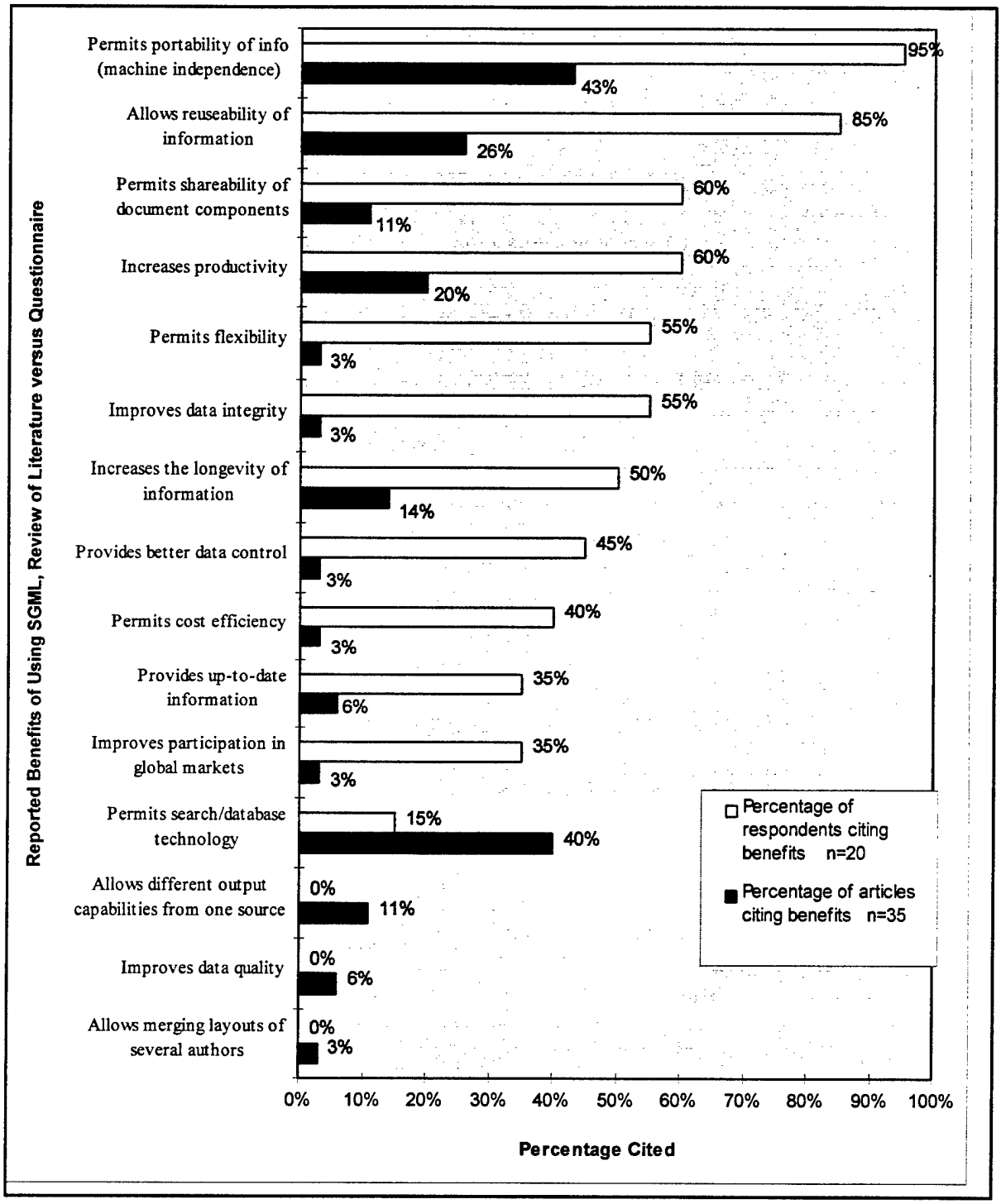

Figure 7: Comparison of Respondents Citing Benefits to Articles Citing Benefits

Figure 7 indicates that the responses from the commercial enterprises confirm and elaborate the information gathered from the literature review. The analysis of this information indicates that the benefits as found in the literature review are consistent with the responses of the commercial enterprises. Portability of information is the most cited benefit of using SGML. Reusability, shareability, increased productivity, improved data integrity, flexibility, and information longevity remain in the top half of the benefits for the commercial enterprises. The one surprising advantage that did not rank in the top 50 percent for the commercial enterprises was database technology. According to the questionnaire responses, it appeared under "other" and ranked second to last. The literature review indicated that the ability to search documents and database technology is emerging as a benefit of using SGML and was cited in 40 percent of the sources collected. This may be an indication that the commercial enterprises are not fully exploiting all of SGML's benefits.

One benefit that did not receive much attention from the sources collected in the literature review and commercial enterprises was security. The Office of the Secretary of the Air Force is concerned about security because there are different classifications and levels of security for information which is limited to authorized individuals. Air Force personnel have the responsibility to ensure that classified information is protected from compromise by proper classification and safeguarding the information. The Air Force does not allow access of classified information to unauthorized individuals or organizations. The migration to SGML provides the capability of on-line publications

and this leads to security risks and a possible compromise of information which needs to be addressed in future research.

Two sources provided information on SGML's capability to handle security issues. These sources indicated that SGML allows for tagging of specific sections of documents which will only allow individuals with proper levels of security access to the appropriate information (Lunemann, 1995 and November 1995, WWW). This security feature is possible with the use of a front-end application that would check the user's user identification and password. The front-end application is a simple software application that could be created by the system administrator.

According to the literature review and the questionnaire responses, the effects of implementing SGML in terms of cost/benefits are difficult to compute. Twenty-five percent of the commercial enterprises have conducted a cost/benefit analysis. The comments provided about their results do not give specific dollar figures for the cost savings. The responses indicate intangible benefits such as data reuse and accessibility to information in multiple output formats. The lack of documented information on cost/benefits of SGML available from the literature review indicates paybacks and return on investment for SGML are hard to compute. Turner commented that it is difficult to measure costs and productivity for document-type data (Turner, 1994: 56). There is no baseline from which to measure the cost/benefits, and the implementation of SGML usually results in significant changes in the way an organization processes its documents, making results harder to measure. Therefore, it is not surprising that only 25 percent of

the respondents have conducted a cost/benefit analysis. The literature review did provide the following as intangible benefits:

- Permits companies to save money by producing documents on-line instead of in hardcopy
- Reduces time and money that would be required to produce and mail hardcopy documents
- Decreases amounts of paperwork
- Provides easy transfer and access of electronic information
- Provides up-to-date information on a timely basis
- Allows compliance to CALS standard or other mandated standards
- Permits ability to reuse data
- Allows migration of documents to a wide range of output specifications

Research Question Four

Research question four asked, "What underlying problems or concerns are associated with these commercial enterprises achieving the benefits of SGML?" Questions 18 and 23 of the questionnaire gathered valuable information to answer this research question. The responses from the commercial enterprises indicate that there are problems that exist in several areas of SGML. The top two problems are in the area of conversion and training.

Converting documents from paper has been identified as a major hindrance in the process of implementing SGML. The respondents' comments indicate that the source (paper-based product) is unstructured and requires manual cleaning up prior to SGML tagging as the main reasons for slowing down the process of converting documents to SGML. According to Gross, traditional sources of data do not have explicit structural information (Gross, 1993: 220). SGML requires an explicit and consistent structure of the document if the conversion is to be completely successful. One respondent indicated

that the Document Type Definition (DTD) used for conversion is often different than the DTD for the target document. This indicates that the document being converted may need to be restructured to work with the target DTD.

Gross recommends development of a pre-conversion plan in order to analyze the degree of difficulty of parts of the document and volume of information to be converted in order to gauge the appropriate blend of manual and automated processing to be applied. The steps to Gross' conversion planning are useful for the Air Force publishing centers to follow (Gross, 1993: 225). They include the following:

- Be selective in what you convert
    - Separate materials that will be reused
    - Determine best sources for each section
- Have the following on hand
    - Hardcopy of the document, if it exists
    - Inventory of electronic sources
    - The DTD (your specification for the conversion)
- Define the approach you will use
    - Determine what is important
    - Separate those parts that are easier to do by hand. Don't try to automate everything
- Implement on a small scale (prototype)
- Plan for a review (conversion will not be 100%)
- Set up conversion to be done in stages

The second most common problem identified by the respondents is SGML training. Training is a problem in that it is often overlooked in an effort to get the product on line immediately. One respondent indicated that training is best provided by users of the language as opposed to instructors. Instructors providing the training may not know the details about the organization's use of SGML and such lack of knowledge only hinders the learning for the individuals within that organization. However, there is

minimal information available in existing literature on SGML training and the problems associated with it. Several articles indicated that SGML is like a programming language, and it takes considerable expertise to implement it. However, none of the sources collected on SGML addressed training issues.

The Air Force Electronic Publishing Master Program Guide (1996) has indicated a need to provide training to various levels of individuals within the Air Force's electronic publication process. The Air Force's plan for addressing the training needs is to ensure that at a minimum one individual per MAJCOM and Air Staff publication personnel receive extensive hands-on training. The comments received from the commercial enterprises about "writer buy-in," "lack of experienced people," and "conversion and authoring require a new mindset" indicate that the Air Force should re-address their training plan to provide additional focus on the individuals at base level who are also required to convert and author publications. If hands-on training is not provided to these individuals, they will not be able to efficiently and effectively do their jobs in the publication process. Another problem within the Air Force is that base level authors are using a DTD based on the Air Staff and MAJCOM structure. Before the base level publication centers can use the DTD effectively it must be tailored to the structure of their target documents. Training on how best to tailor the DTD would be beneficial to base level publishing personnel

There are several areas besides conversion and training associated with SGML implementation that are not fully developed These problem areas include authoring, tagging of documents, writer buy-in, lack of experience with SGML, top management

support, and SGML's complicated style. With the increasing number of products and tools available for organizations to use for conversion and authoring in SGML, many of these problems may go away soon. For the Air Force to fully exploit the benefits of SGML, it must understand and implement SGML in a way that takes full advantage of these benefits. As indicated from the literature review and the questionnaire, there are many valuable sources available for information on SGML in journals, books, and the WWW. Training for users of SGML would increase their knowledge of SGML and provide insight into many benefits they may not be aware of.

Research Question Five

"Does SGML training affect an organization's use of SGML?" is the fifth research question. To answer this question, the commercial enterprises were asked to respond to two questions pertaining to their organization's training efforts and training plans.

The responses from the questionnaire indicate that sixty percent of the commercial enterprises provide training to their employees, and all feel that it has enhanced their employees' use of SGML. Thirty-five percent do not provide training and of these 35 percent, 85 percent believe training would be beneficial to enhancing their employees use of SGML.

Forty-five percent of the respondents have a training plan for the conversion/authoring of SGML documents and 44 percent feel the training plan is "very effective," while 55 percent said it "meets the needs of the organization." Fifty percent of

the commercial enterprises do not have a training plan and of these enterprises, 80 percent believe one would be beneficial. Ten percent of the organizations that do not have a training plan said it would not be beneficial and one respondent was not sure if a training plan would help.

The commercial enterprises have indicated that SGML training is essential in improving the employees' use of SGML. The 35 percent of the organizations that have not provided training and the 50 percent that have no training plan, these percentages suggest the need for training issues to be addressed. None of the sources collected for the literature review on SGML discussed training. Since training is an important aspect of learning and there are several elements of SGML that are complicated and difficult to master, SGML training issues should be included in future research on SGML.

Research Question Six

Research question six asked, "Can a commercial enterprise be used as a benchmark for the Air Force in its migration to an electronic publishing environment?" In order to answer this question, two types of research were completed. The first was a literature review to determine the criteria for benchmarking. The second was gathering data from commercial enterprises.

The literature review on benchmarking indicated that there are several models available to guide prospective benchmarkers. The information collected and the number of references to Xerox's 10-step process suggests that this model is the simplest and outlines a process that the Air Force could follow to benchmark a commercial enterprise.

Forty percent of the journal articles indicated that a company's direct competitors are not always likely to know the best way to perform a function or process. Rather than comparing themselves to direct competitors in the same industry, organizations can examine practices and methods of those organizations in dissimilar industries, which is called functional benchmarking. Studying dissimilar industries allows the benchmarking organization to find industry-wide best practices *not* the best practices within a particular business sector. The journal literature suggests that the type of benchmarking that best fits the need of the Air Force for this study is functional benchmarking. Functional benchmarking requires the comparison of functions or processes. This type of benchmarking involves comparisons against functional competitors or industry leaders, even if they are in dissimilar industries. Functional benchmarking would involve, in the case of SGML, those firms that are recognized as having superior functions in the area of conversion or authoring of SGML documents wherever they exist.

Robert Camp states that functional benchmarking makes investigation and sharing of data easier (Camp, 1989: 64). There is a natural interest among competing organizations to understand similar practices and see how they compare. The commercial enterprises responding to this study's questionnaire were willing to share information on their use of benefits from and problems associated with their implementation of SGML. On the other hand, they also asked if the information gathered and analyzed could be provided to them so they could compare the other industries' responses to theirs.

As addressed by the sources collected, the benefits of benchmarking outweigh the drawbacks. The drawbacks are mainly management issues that can be resolved over

time. The principal benefit mentioned by all the sources was enabling the best practices from any industry to be incorporated into the benchmarked function. If company X can capitalize on another's best practices, it can leap ahead of the competition or become the best at what it does.

The process of conversion and authoring of publications with SGML can be benchmarked as long as the Air Force successfully determines which organization to benchmark. Although benchmarking requires using the best industry practices, no ranking of electronic publishing experts using SGML existed in the literature. Several questions in the questionnaire were developed to gather information to determine if a commercial enterprise could be used as a benchmark for the Air Force. The first five questions were asked to solicit demographic information on the commercial enterprises. This demographic data provided information on the type of organization, number of personnel, and how many people are using SGML. This data can then be compared with the Air Force's demographics to see if a similar commercial organization exists that can be selected as a benchmark.

The demographic data collected indicates that several enterprises are similar to the Air Force in terms of size. In comparison to the number of Air Force personnel involved with the base level publication process, these commercial enterprises are similar in size and the number of employees and contractors involved with SGML. More than 50 percent of the commercial enterprises employ more than 1000 employees, indicating that the commercial enterprises are large organizations like the Air Force. Further

demographic data indicated that more than 25 employees are using SGML and a mean of 2.5 contractors are involved with the SGML process.

Ten questions in the questionnaire were developed to gather information on the commercial enterprises' use of SGML. The responses from the commercial enterprises provided information that is useful in determining if their reasons for using SGML are similar to the Air Force reasons. The Air Force was driven to SGML mainly by changes in technology and customer demand. The only way to provide Air Force customers with up-to-date information they require is by using current technology. The responses from the commercial enterprises indicate that customer demand and a change in information technology are also drivers for their conversion to or implementation of SGML.

To elicit responses which would indicate the experience level of each organization using SGML, the commercial enterprises were asked the number of pages converted to SGML and the number of SGML documents authored per year. The responses indicate that 90 percent of the organizations have converted more than 2000 pages of information, and 65 percent of the organizations author 25 or more documents per year. These responses suggest that several commercial enterprises could provide useful information to the Air Force publishing centers on their methods and techniques which have enable them to convert and author documents using SGML.

The Air Force has adopted SGML for its electronic publications. The commercial enterprises have indicated that their top two uses of SGML are for electronic books and on-line distribution. However, the distribution of responses indicates no one single use for SGML that stands out, but many uses and reasons behind the commercial enterprises'

decisions to implement SGML. These commercial enterprises may have converted to SGML originally for one purpose but after the conversion adopted other uses as well.

Two questions in the questionnaire were asked to elicit responses on the type of documents and information contained within the documents. The responses indicate that SGML is used primarily for technical manuals and user manuals. The distribution of responses across all types of information indicate that the commercial enterprises are taking full advantage of SGML and using it for more than just standard text. In fact, virtually all types of information including figures, graphics, links, tables, and mathematical equations are being incorporated into the documents by these commercial enterprises.

The questionnaire also included four questions regarding source documents, types of authoring, types of operating systems being used, and selection of products/tools. These questions were asked to determine which commercial enterprises could provide information to the Air Force on their experience with specific SGML applications, SGML products/tools, and SGML implementation techniques. Since responses show that more than 60 percent of conversion is from legacy and electronic sources, they would be able to share valuable information and insight into the best conversion techniques with Air Force publishers. Since more than 70 percent of the commercial enterprises stated they use direct authoring, they have probably created templates or DTD's to meet their authoring needs. The development of these techniques by the commercial enterprises would provide the Air Force publishers with information on how best to develop a DTD for Air Force publications at various levels within the organization.

Eighty percent of the respondents indicated that they use MS-Windows; 55 percent indicated they use MS-DOS. These are the same environments that the Air Force publications centers operate in. Since there is a similarity in these operating environments, the commercial enterprises could provide valuable information to the Air Force if they were to be benchmarked.

The responses to the question which asked about product/tools used for conversion and authoring in SGML indicate that there are many resources available for use. These responses provided useful information on products/tools that the Air Force could use for its migration to SGML. The Air Force has already selected FastTag for its conversion efforts and InContext for editing and authoring. However, Air Force authors and editors in the field are welcome to use any tool they wish as long as they produce SGML documents compliant to the DTD (Department of the Air Force, 1996). Forty percent of the respondents are currently using FastTag and 15 percent are using InContext. The Air Force publishing centers could contact these organizations to ask questions about problems or request clarification of techniques. However, if the Air Force authors and editors wish to select another tool, the above responses indicate there are many organizations that have selected various products/tools are used on a daily basis to convert and author documents.

Based on the information gathered from the literature review and the questionnaire, the Air Force could benchmark a commercial enterprise to improve or streamline the Air Force's migration to an electronic publishing environment. The commercial enterprises must work hard to stay ahead of their competition. Many of the

commercial organizations that participated in this study have been using SGML for more than seven years and have the expertise which would be useful to the Air Force as it starts its migration toward an electronic publishing environment. The Air Force should take advantage of the information available from these organizations and learn from them.

Conclusions

The analysis of the information gathered for this study indicates that SGML is a powerful tool that the Air Force electronic publishing program should use to improve the efficiency and effectiveness of its publication process. SGML provides a standardized approach that would allow the publishing centers to reduce the time it takes to publish Air Force documents, standardize document structure, eliminate storage of paper-based products, provide up-to-date and accurate information, and permit point-to-point availability of publications electronically on personal computers.

The management of documents -- creating, distributing, retrieving and reusing -- used to be hampered and prevented by obstacles imposed by differing hardware and software architectures, operating systems, and application-specific file formats. With the use of SGML, many industries are recognizing benefits to overcome these obstacles. Research has indicated that the four most cited benefits are portability of information, reusability, shareability, and increased productivity. However, many other benefits are also possible with the use of SGML, and the Air Force publishing process should take the necessary steps to reap all of SGML's benefits.

SGML is beginning to play an increasing role in information services available through the WWW and Internet. Publishing information on the WWW offers the opportunity to satisfy more customers and provides customers with just-in-time access to information around the clock. The WWW can provide the Air Force publication centers and end users widespread visibility of information because the WWW stretches across the globe and across boundaries in computer hardware and software. The use of the WWW facilitates universal access to unrestricted standard publications and forms that Air Force personnel use on a daily basis. This universal access reduces hurdles normally faced with hardcopy distribution, such as high production costs and inaccurate information.

Distributing information via the WWW requires publishing information in only one place. Once published, the information is available to any of the million web users. Since the WWW provides a central repository for electronic document distribution, it simplifies the document update procedures and version control. Air Force publication managers simply replace the old version on the web with the new release and anyone accessing the documents gets the most current information. The WWW also provides users with the capability to query documents and search for information or related topics they request. Retrieval is done in a matter of seconds, compared to days when requested by hardcopy from the publication distribution office. The text search and retrieval capability is accomplished through links to various, interrelated documents. This means that Air Force publications can be linked to each other and to other required documents, such as forms, that are referenced by them. This makes the process of accessing a publication with a supplement very simple. Since the documents would be linked, the

supplemented information would appear in the requested publication and there is no need to access two separate documents.

SGML is an enabling technology that is a means to an end, not a solution in itself. It is not easy to implement SGML. The results of the questionnaire indicate several areas of concern for the commercial enterprises -- SGML training, conversion, authoring, and tagging of documents. The commercial enterprises have indicated that once a decision is made to adopt SGML, they have looked toward commercially available SGML tools to help them achieve their information needs. Programming tools and sophisticated editors and parsers are commercially available for SGML users. The tools that are selected for conversion and authoring of SGML documents should meet the needs of the organization using them. There is an abundance of information about products and tools available on the WWW and Internet. However, there is also valuable information from commercial enterprises that we can tap into.

Commercial organizations must work smart to stay in business and ahead of the competition. The Air Force can learn a lot from them with the use of benchmarking. Since effective benchmarking does not depend on following practices from within the same industry, the Air Force could benefit by augmenting the best industry practices and techniques associated with the conversion and authoring of documents using SGML from a commercial enterprise. Functional benchmarking of dissimilar industries needs to be carefully planned. A solid strategy would allow the Air Force to examine the practices of commercial enterprises so that Air Force publication centers can integrate the useful pieces of information into their electronic publication process.

Recommendations

On the basis of the research performed and reported in this study, the following recommendations are made:

1. **The need for SGML training**. The commercial enterprises have indicated that training enhances an employee's use of SGML. The Air Force's Electronic Publishing Master Program Guide states that training should be provided to a minimum of one individual per MAJCOM and Air Staff publication personnel should receive extensive hands-on training. The comments from the commercial enterprises indicate that training needs to be focused on the individuals who author and convert documents to SGML. The Air Force planners should review their guide and include extensive hands-on training to base level publication personnel so that they have a complete understanding of SGML's capabilities and benefits. This training will allow the authors and editors of publications to do their job in an efficient and effective manner.
2. **Development of a Document Type Definition (DTD) for base level publication centers**. A DTD describes the structure of a document and functions as a template to identify the type of documents and elements of the document in the order in which they appear. The Air Force planners have developed a DTD for Air Force publications based on the Air Staff and MAJCOM structure. Since the DTD defines all elements that are allowable and common to a given type of document, this indicates that the base level publication centers are using a DTD that is very large and encompasses elements that are not specific to their needs. The base level publication

managers should review the Air Force level DTD and redefine it to fit their needs. Several commercial enterprises have provided comments concerning the size of their DTD and they can provide valuable information concerning the best technique(s) to apply.

3. **Use of the WWW as a central repository**. The WWW has proved to be a useful tool in distributing information to millions of web users across boundaries in computer hardware and software. The Air Force electronic publication process can benefit from the use of the WWW by providing universal access of unrestricted publications and forms to personnel who require them on daily basis. Air Force publication planners should consider the WWW as its central repository for electronic documents. This repository would simplify document update procedures and version control, while providing widest dissemination of information to personnel in a timely manner.

4. **Benchmarking commercial enterprises**. Benchmarking is a valuable process that the Air Force should use to study and learn about SGML from commercial enterprises. As indicated earlier, the commercial enterprises that participated in this study were more than willing to provide information about their organization's use of SGML. The Air Force publication planners can use functional benchmarking to obtain useful information and integrate those bits and pieces of useful information into their electronic publication process. Questions that could be asked of the commercial enterprises during the benchmarking process include the following:

    - How has your organization implemented SGML?

- What have you learned from your implementation efforts?
- What solutions, products, or tools have you used?
- How did you start the process?
- What were the expenses?
- Who should be trained and what is your training plan?
- How does SGML affect authors and end-users?
- How are SGML applications maintained?
- How have you taken advantage of SGML-coded documents?
- What technologies are available and which have you used?

5. **Establishment of a systematic implementation process.** The Electronic Publishing Master Program Guide provides a starting point from which the process can be developed. Several areas that are not clearly addressed in the guide are the development of base level procedures for conversion and authoring of SGML documents, and establishment of an Air Force-wide central repository for electronic publications. The base level procedures can be developed with input from individuals experienced with using SGML, base level publication personnel, and commercial enterprises. Several commercial enterprises have been using SGML to create and edit documents for more than seven years. These organizations can provide insight into issues that may not have been considered by the Air Force publication planners. In addition, a carefully developed strategy should be developed that addresses the establishment of an on-line central repository. This strategy should include the needs of the Air Force, the publications which are important, access control considerations, and a survey of techniques or standards which support the central repository.

Limitations of the Study

As with any research study, this study was not without limiting factors. There were several limiting factors that affected how this research was conducted and which affected the data collected. However, none of these limitations should significantly impact the study's validity. This study is subject to the following limitations:

1. **Qualitative Nature of the Study**: As described in Chapter Three, this research is an exploratory, qualitative study. As such, this research provides no results of statistical significance. In the researcher's view, the scope and objective of the study do not lend toward a quantitative approach to the research problem.
2. **Sample Size**: The sample size for this research is 20, which may be considered too small to adequately represent commercial enterprises who have converted documents to SGML or have been using SGML within their organization to author documents. Sample size was not deemed as important as industry-wide representation by organizations. Furthermore, the selection criteria for the commercial enterprises was based on the organization's use of SGML. The researcher believes that the quality of the sample is more important to this study than the quantity of the sample.
3. **Longitudinal Study**: Time constraints in collecting the necessary information from the commercial enterprises limited the data collection methods that could be used. A longitudinal study of the commercial enterprises would have provided the means to track the changes of the commercial enterprises' use of SGML and identified the key benefits each organization has realized over time. In addition, a longitudinal study

could have provided the researcher with the opportunity to study and monitor a commercial enterprise's benchmarking processes and the benefits associated with its implementation.

4. **List of Organizations to Benchmark**: Although benchmarking requires using the best industry practices, there was no ranking or list of commercial enterprises using SGML. The researcher developed a population of 65 commercial enterprises based on information gather from existing literature and the WWW. This representation was deemed appropriate since the most experienced or knowledgeable person involved with the organization's migration and implementation of SGML provided a reply to the questionnaire.

Future Research

Several opportunities exist to study other areas related to this research effort. The following are areas recommended for future study:

1. **Access to the World Wide Web using SGML documents.** Leading vendors in the SGML industry are exploiting their technology for Internet applications that will change the way organizations do business (White Paper 4002-II, 1995: 6). Air Force publications can be converted to a format which allows the documents to be placed on the WWW for on-line access to military members. However, the large SGML documents do not work well with the existing WWW browsers. Research on the techniques and strategies to convert these large SGML documents to WWW documents should be accomplished in order to keep up with the rapid development

of WWW and Internet applications. The Air Force electronic publishing process currently focuses on the use of SGML while the WWW and Internet focus on HTML. Specific research should be directed to the use of SGML versus HTML and a conversion technique that allows the WWW or Internet to read SGML *and* HTML documents.

2. **Implications of Information Warfare for an electronic publishing environment.** The use of the WWW facilitates universal access to unrestricted standard publications and forms (and many other types of information) that Air Force personnel use on a daily basis. This new technology can provide Air Force members with information at their fingertips, but it also can have underlying implications. "The United States is leading the world into a globally networked society, a true Information Age where information and economic value become nearly synonymous" (Schwartau, 1994: 12). While we are leading the evolution of a networked society, we should also be concerned and wary of the vulnerabilities associated with such new technologies. Schwartau says that information is intangible and does not have an immediately quantifiable monetary worth unless you lose it or someone manipulates it to suit his particular goals. For this reason, it is important that we take a closer look at the implications of information warfare as we move toward a paperless Air Force. More research in the area of encryption schemes and risks associated with on-line access to information needs to be conducted.
3. **SGML training and its effect on authors and users of SGML.** Responses to the questionnaire state that SGML training is essential in improving the employees' use

of SGML. Since 35 percent of the organizations have not provided training and 50 percent do not have a training plan, training issues should be addressed. The lack of information in the literature review on SGML training reinforces this recommendation.

4. **Identification of commercial enterprises to benchmark.** The process of conversion and authoring of publications with SGML can be benchmarked as long as the Air Force can determine which organizations to benchmark. Although benchmarking requires using the best industry practices, no such ranking of electronic publishing experts using SGML exists in the literature. Additional research in this area would be useful.

Summary

This thesis addressed the benefits of SGML that would enable the Air Force to enhance future publications capabilities. It determined that benchmarking can improve and streamline the Air Force's migration to an electronic publishing environment. The commercial enterprises participating in this study indicated that the benefits of implementing SGML outweigh the costs in terms of flexibility and accessibility of information in documents that are offered to the customer and producer. In addition, SGML allows portability of information, reuseability, shareability, and information longevity. SGML is an enabling technology that is a means to an end, not a solution in itself. Only through carefully planning and training will the Air Force be able to fully exploit the benefits of SGML. Benchmarking can provide an avenue to do this. By

implementing benchmarking, the Air Force could examine the best practices of commercial enterprises and implement changes based on their observations.

Appendix A: Glossary of Terms

ASCII - American Standard Code for Information Interchange

Air Force Catalogs - informational publications that are detailed listings which describe or list a collection of information according to some plan (AFI 37-160, Vol 1: 10).

Air Force Directories - are informational publications that are compilations serving to direct, that are systematically arranged, usually in alphabetical or classed order, such as listings of addresses, affiliations, functions, and similar data (AFI 37-160, Vol 1: 10).

Air Force Doctrine Documents (AFDD) - Contain broad, enduring guidance for preparing and employing United States Air Force resources to support National objectives (AFI 37-160, Vol 1: 9).

Air Force Handbooks (AFH) - concise reference books on the technical aspects of a particular subject or a compilation of factual data and instructional material not subject to frequent revision (AFI 37-160, Vol 1: 10).

Air Force Indexes - are informational publications that serve to guide, point out, or otherwise facilitate reference (AFI 37-160, Vol 1: 10).

Air Force Instructions (AFI) - are orders of the Secretary of the Air Force and are approved in the Secretariat or the Air Staff and will be issued as departmental publications (AFI 37-160, Vol 1: 9).

Air Force Manuals (AFMAN) - are guidance documents consisting of procedures that usually contain examples for performing standard tasks, supporting education and training programs, or computer system operating instructions (AFI 37-160, Vol 1: 9).

Air Force Mission Directives (AFMD) - prescribe the mission, area or responsibility, organization, responsibilities, and command relationships of Air Force units (AFI 37-160, Vol 1: 9).

Air Force Operating Instructions (OI) - assign responsibilities, direct actions, and prescribe procedures within a headquarters, organizational element, or functional area (AFI 37-160, Vol 1: 9).

Air Force Official Bulletins - contain temporary announcements, notices, and instructions (AFI 37-160, Vol 1: 10).

Air Force Pamphlets (AFP) - informational publications which normally are "how to" documents and include procedures for implementing Air Force policies (AFI 37-160, Vol 1: 10).

Air Force Policy Directives (AFPD) - are orders of the Secretary of the Air Force and contain directive policy statements of the Office of the Secretary of the Air Force and Headquarters United States Air Force to initiate, govern, or regulate actions of the conduct, within their specific areas of responsibility, by the Air Force activities at any level (AFI 37-160, Vol 1: 9).

Air Force Publishing Bulletins (PB) - announce processing and rescission actions on Air Force publications and forms (AFI 37-160, Vol 1: 10).

Air Force Staff Digests - contain summaries of significant staff actions, important announcements, and special notices (AFI 37-160, Vol 1: 10).

Air Force Supplements (AFS) - add material to publications issued by higher headquarters or agencies (AFI 37-160, Vol 1: 9).

Air Force Visual Aids (VA) - they can be posters or graphic illustrations (AFI 37-160, Vol 1: 10).

ASCII (American Standard Code for Information interchange) - A 7-bit standard code used for information interchange among data processing systems, communications systems, and associated equipment. This standard character encoding scheme is used extensively in data transmission. (Almost all publishing systems will accept ASCII.)

Benchmark - a continuous process of measuring an organization's products, services, and practices against other firms recognized as industry leaders (Camp, 1989: 248).

Business Processes - a collection of activities that takes one or more kinds of inputs and creates an output that is of value to the customer (Hammer, 1993: 35).

Continuous Acquisition and Life-cycle Support (CALS) - CALS began as primarily a US defense industry and government effort to integrate systems development, production and support, but has become recognized as a leading-edge prototype for a manufacturing community. CALS was developed to control the mounds of paperwork generated with the design, development, and manufacturing of weapons systems. For the purpose of this research the word CALS will be

defined as the following function: the use of digitized technical data in lieu of hard copy media in an intelligent or knowledge enhancing manner (Elliot, 1993: 47).

Document Type Definition (DTD) - A DTD is a formal definition of the elements, structures, and rules for marking up a given type of SGML document. You can store a DTD at the beginning of the document or externally in a separate file (October 1995, WWW).

HyperText Markup Language (HTML) - a subset of SGML. HTML is based on SGML, with its own Document Type Definition and a fixed tag set. HTML markup provides text searches and retrieval capabilities through links to various, interrelated documents.

Information System - an organized set of formal systems for obtaining, processing, and delivering information in support of the business operations.

Information Technology - computing hardware and software that can be used to create new business opportunities.

International Organization for Standardization (ISO) - The ISO is an industry-supported organization that establishes world-wide standards for everything from data interchange formats to film speed specifications (October 1995, WWW).

Markup - Markup is anything added to the content of the document that describes the text (October 1995, WWW).

Open Systems Architecture - software systems that can operate on different hardware platforms because they are built on non-proprietary operating systems, user interfaces, application standards, and networking protocols (Lauden, 1996).

Standard Generalized Markup Language (SGML) - SGML, also know as ISO 8879, is an international standard for document description published by the International Organization for Standardization and was released in 1986. SGML enables a flexible interchange of a document's content across heterogenous systems.

Tags - SGML tags are inserted around the content in order to create an SGML document. Tags consist of pieces of code surrounded by the symbols "<" and ">" and are used to indicate structural information or formatting.

Appendix B: Commercial Enterprise Mailing List

Name: Silvia Sirotich
Organization: Active Systems Inc

Name: Joe Beckner
Organization: Hoechst Marion Roussel, Inc

Name: Jasper Rose
Organization: Active Information Management

Name: Glenn Handrahan
Organization: Advanced Engineering and Research Assoc, Inc.

Name: DeJean Melancon
Organization: Advantage KBS, Inc.

Name: Glenn Hadrahan
Organization: AERA, Inc

Name: Cornell University
Organization: Albert R. Mann Library

Name: CIO
Organization: Alpnet

Name: Robert Fye
Organization: Aquidneck Management Associates

Name: Tommie Usdin
Organization: ATLIS Consulting Group

Name: CIO
Organization: AutoDesk, Inc.

Name: CIO
Organization: Borland International, Inc.

Name: Theodora Landgrem
Organization: BTS

Name: Dennis J. OConnor
Organization: Bureau Of National Affairs, Inc

Name: Dennis Arnon
Organization: CAL SGML Users' Group

Name: Michael Tobin
Organization: Cambridge Technology Partners

Name: Alison Macintosh
Organization: Canadian Standards Assoc

Name: Ray Henry
Organization: Communitec, Inc.

Name: Richard Oswald
Organization: Control Data Systems, Inc.

Name: Tom Edgerton
Organization: Conversion Data Services, Inc

Name: Gregory Matthews
Organization: Cray Research, Inc.

Name: Mark Gross
Organization: Data Conversion Laboratory

Name: Michael Mercer
Organization: Deere & Co

Name: Bill Cap
Organization: Diebold, Inc.

Name: Alan Hobgood
Organization: Docucon

Name: Sharon Adler
Organization: Electronic Book Technologies

Name: Kevin Rees
Organization: Enterprise Translations

Name: Ed Brackus
Organization: EWB & Associates, Inc.

Name: Eric Skinner
Organization: Exoterica Corporation

Name: Ray Ansley
Organization: Gateway Technologies

Name: John Bowers
Organization: GE Aircraft Engines

Name: Marion Elledge
Organization: Graphic Communications Association

Name: CIO
Organization: Grolier Interactive, Inc.

Name: Hank Pelletier
Organization: GTE Government Systems

Name: Katherine Armstrong
Organization: Hewlett-Packard

Name: Daniel Chang
Organization: HYNET Technologies

Name: Andrea Otken-Dennis
Organization: Imonics Corporation

Name: Clive C. Carpi
Organization: InfoDesign Corporation

Name: Paul Blumfield
Organization: Information Mapping, Inc.

Name: Charles F. Goldfarb
Organization: Information Mgt Consulting

Name: David Silverman
Organization: Innodata

Name: Bryan Layton
Organization: Intergraph Corp

Name: CIO
Organization: International Thompson Publishing

Name: Doug Welling
Organization: JANA Inc

Name: Ginger Stack
Organization: Jouve Data Management

Name: Holly Smith
Organization: Lexicon Systems, Inc

Name: Chet Ensign
Organization: Logical Design Solutions, Inc

Name: Linda Wolpert
Organization: Lucent Technologies

Name: CIO
Organization: Matthew Bender Law & Tax Publisher

Name: Larry Jackson
Organization: Motorola, GSTG

Name: CIO
Organization: New Yorker Magazine

Name: Larry McKinley
Organization: O'Neil & Assoc, CALS Coordinator

Name: Dr. James Mason
Organization: Oak Ridge National Library

Name: CIO
Organization: Ontario Hydro Electrical

Name: CIO
Organization: Oracle Corporation

Name: Editorial Dept
Organization: Oxford University Press

Name: Vance Nakamoto
Organization: Passage Systems

Name: Tracy Oltmann
Organization: PRC, Inc.

Name: Roland Brooks
Organization: Raytheon Service Co

Name: Nacia Avera
Organization: Semiconductor Research Corporation

Name: Janice M. McNulty
Organization: SGML Solutions

Name: Edward T. Kell
Organization: Sikorsky Aircraft, Inc.

Name: CIO
Organization: Storage Technology Corp.

Name: CIO
Organization: Sun Microsystems, Inc

Name: CIO
Organization: Sybase, Inc.

Appendix C: SGML Questionnaire

Shari T. Miles, Captain, USAF
Graduate Student, School of Logistics and Acquisition Management
Air Force Institute of Technology, AFIT/LAR
2950 P Street
Wright-Patterson AFB, OH 45433-7765

Name and Address

Dear ------

I am a student at the Air Force Institute of Technology (AFIT) pursuing my masters degree in Information Resources Management. My thesis research, part of the degree requirement, involves the capabilities and use of SGML. More specifically, I am studying commercial enterprises that have already converted or are using SGML in their day-to-day business activities. The attached questionnaire will provide information to help predict the impact of the conversion of Air Force publications/regulations to SGML and how the Air Force can exploit SGML's capabilities.

I obtained your name and address from the SGML Newswire Hit List as an individual/organization who may be able to provide helpful information on SGML. If you are willing to participate in my thesis research, the following questionnaire asks several questions about you or your organization's involvement with SGML.

Please take the time to answer all of the questions thoroughly. My goal is to receive all the completed questionnaires by 30 May 1996. I will analyze the results and provide a report of the findings back to you within two weeks. Please return the completed questionnaires in the return addressed envelope. You may also fax, or e-mail your response to the following address:

Shari T. Miles, Capt, USAF
Student, Information Resources Management
AFIT/LAR, 2950 P Street
Wright-Patterson AFB OH 45433-7765

Internet: smiles@afit.af.mil
Voice: 513-255-7777 (785-7777 DSN) Ext. 2255
Fax: 513-476-7988 (986-7988 DSN)

Thank you in advance for your time and participation. If you have any questions please feel free to contact me at the above address.

Sincerely

Shari T. Miles, Capt, USAF
Graduate Student, Information Resources Management
Graduate School of Logistics and Acquisition Management

2 Attachments:
1. SGML Questionnaire
2. Return Envelope

The Air Force is migrating from paper to electronic publishing. To do this properly the Air Force needs to consider the electronic publishing business from the headquarter's level to base level, from the author to end user. This migration suggests the need for a comprehensive standardized approach. To this end, the Air Force has selected SGML to implement Air Force wide. The information gathered from this questionnaire will be used to predict the impact of the conversion of Air Force publications/regulations to SGML and how the Air Force can exploit SGML's capabilities.

Name: _____ Duty Title: _____

Organization Name: _____

Address: _____

Phone Number: _____ E-Mail Address: _____

Please answer the following questions by marking ALL the boxes that apply:

1. What is the primary activity of your organization?

❑ Business/industry ❑ Computing ❑ Government ❑ Education
❑ Research ❑ Publishing ❑ Communications ❑ Services
❑ Public Admin ❑ Manufacturing ❑ Retail ❑ Wholesale
❑ Other
(please specify)_____

2. How many employees are in your organization?

❑ 0 to 25 ❑ 26 to 250 ❑ 251 to 500 ❑ 501 to 1000 ❑ More than 1000

3. How long has your organization been using SGML?

❑ Less than 1 year ❑ 1 to 2 years ❑ 2 to 3 years ❑ 3 to 5 years
❑ 5 to 7 years ❑ More than 7 years

4. How many people in your organization are using SGML? _____

5. How many outsourcing contractors in your organization are using SGML? _____

6. Why have you chosen to use SGML?

❑ Required by upper management
❑ Compliance to a standard or contractual agreement
❑ Other (please specify) _____
❑ Required to meet demand of customers
❑ Competition
❑ Change in information technology

7. How many pages have you converted to SGML?

❑ None ❑ 1 to 100 ❑ 101 to 500 ❑ 501 to 1000
❑ 1001 to 2000 ❑ More than 2000

8. How many SGML documents do you author per year?

❑ None ❑ 1 to 10 ❑ 11 to 24 ❑ 25 or more

9. For which of the following do you use SGML?

❑ On-line distribution ❑ Storage ❑ Document management
❑ Electronic books ❑ CD-ROM ❑ Intent for World Wide Web use
❑ Legacy documents ❑ Information preservation ❑ HTML/World Wide Web
❑ Varied platform use ❑ Ensuring structural consistency
❑ Other (please specify) _____

10. For what types of documents are you using SGML?

❑ Letters/Memos ❑ Books/Publishing ❑ Electronic Data Books
❑ User Manuals ❑ Regulations/Publications ❑ On-line Electronic Database
❑ Technical Manuals ❑ Request for Proposals
❑ Other (please specify) _____

11. What types of information are contained in your documents?

❑ Standard text ❑ Tabular information ❑ Figures and graphics
❑ Math and equations ❑ Links and cross references
❑ Other (please specify) _____

12. Are your source documents ❑ Hardcopy (paper) or ❑ Electronic?

13. What percent of authoring is done with the following methods?

____ Direct authoring
____ Keyboarding from a manuscript
____ Translating from an unstructured electronic source
____ Generating from a database or spreadsheet
____ Scanning from a printed source using OCR
____ Other (please specify)_____

14. What operating systems do you use?

❑ MS-DOS ❑ IBM CMS/MVS ❑ Macintosh
❑ UNIX ❑ MS-Windows ❑ OS/2
❑ Other
(please specify) _____

15. Which of the following products/tools do you using for conversion/authoring?

❑ FastTag ❑ InContext ❑ MS Author ❑ DynaText
❑ Intellitag ❑ SGML Hammer ❑ SGML/Search ❑ WorldView
❑ HyMinder ❑ FrameBuilder ❑ Interleaf ❑ IBM SGML Translator
❑ Write-It ❑ SGML Tagger ❑ Author/Editor ❑ Application Builder
❑ In-house applications
(please specify) _____
❑ Other
(please specify) _____

16. What do you believe are the benefits of using SGML?

❑ Increased productivity
❑ Better data control
❑ Portability of information
❑ Cost efficiency
❑ Information longevity
❑ Security
❑ Other
(please specify)_____
❑ Reusability
❑ Shareability
❑ Flexibility
❑ Up-to-date information
❑ Improved data integrity
❑ Ability to participate in global markets

17. Has your organization conducted a cost/benefit analysis of SGML in general?

❑ Yes ❑ No

If yes, what were the results: _____
_____
_____

18. Which of the following have been problems you've encountered using SGML?

❑ Conversion ❑ Authoring ❑ Tagging of documents
❑ Training ❑ Other (please specify) _____

Please explain why: _____
_____
_____

19. Does your organization provide SGML training to its employees?

❑ Yes. If so, do you feel this training has enhanced employees' use of SGML? ❑ Yes ❑ No

❑ No. If not, do you believe training would be beneficial? ❑ Yes ❑ No

20. Does your organization have a training plan for conversion/authoring SGML documents? (For example: Basic SGML training for new employees, in-depth training after 3 months on the job, continuing training after 6 months on the job)

❑Yes

If yes, how effective is this plan?

❑ Very effective ❑ Meets needs of organization ❑ Not effective

❑ No

If not, would a training plan be beneficial for your organization?

❑ Yes ❑ No

21. What resources have you used to learn about SGML?

❑ Practice/Experience ❑ Books ❑ Newsletters ❑ Training
❑ Co-workers ❑ User groups ❑ Internet
❑ Other
(please specify) _____

22. I may wish to contact you for further information regarding this questionnaire or SGML. May I call you?

❑ Yes ❑ No

23. If there is any additional information you would like to provide?

Thank you for your participation. Please return the completed questionnaire to the following address:

Shari T. Miles, Capt, USAF
Graduate Student, Information Resources Management
AFIT/LAR, 2950 P Street
Wright-Patterson AFB, OH 45433-7765

Internet: smiles@afit.af.mil

Fax: 513-476-7988

Appendix D: Commercial Enterprises that Responded to Questionnaire

Name: Glenn Hadrahan
Duty Title: Electronic Documentation Mgr
Organization: AERA, Inc
Street: 1919 South Eads St
City: Arlington, VA 22202
Phone: 703-486-1993
E-mail: glennh@aera.com

Name: Michael Tobin
Duty Title: Vp Operations, Central Europe
Organization: Cambridge Technology Partners
Street: Apollo House, Appollolaan 15
City: Amsterdam, Netherlands
Phone: 31.20.575.0407
E-mail: mtobi@ctp.com

Name: Alison Macintosh
Duty Title: Mgr, Electronic Publishing
Organization: Canadian Standards Assoc
Street: 178 Rexdale Blvd
City: Etobicoke, Ontario M9W 1R3
Phone: 416-747-2466
E-mail: macintoa@csa.ca

Name: Richard Oswald
Duty Title: Mgr, Documentation Services
Organization: Control Data Systems, Inc.
Street: 4201 N. Lexington Ave.
City: Arden Hills, MN 55126
Phone: 612-482-3302
E-mail: richard.l.oswald@cdc.com

Name: Tom Edgerton
Duty Title: President/CEO
Organization: Conversion Data Services, Inc
Street: 975 Walnut St. Suite 209
City: Cary, NC 27519
Phone: 919-319-1918
E-mail: tedger5944@aol.com

Name: Gregory Matthews
Duty Title: Publications Manager
Organization: Cray Research, Inc.
Street: 655 Lone Oak Drive
City: Eagan, MN 55121
Phone: 612-683-3814
E-mail: gregor@cray.com

Name: Mark Gross
Duty Title: President
Organization: Data Conversion Laboratory
Street: 184-13 Horace Harding Expressway
City: Fresh Meadows, NY 11365
Phone: 718-357-8700
E-mail: markgross@dclab.com

Name: Michael Mercer
Duty Title: Information Technology Specialist
Organization: Deere & Co
Street: John Deere Road
City: Moline, IL 61265
Phone: 309-765-4136
E-mail: mm46100@deere.com

Name: Andrea Otken-Dennis
Duty Title: Marketing Associate
Organization: Imonics Corporation
Street: 111 Corning Dr.
City: Cary, NC 27511
Phone: 919-461-6000
E-mail: adennis@imonics.com

Name: Clive C. Carpi
Duty Title: VP, US Operations
Organization: InfoDesign Corporation
Street: 7700 Leesburg Pike, Ste 204
City: Falls Church, VA 22043
Phone: 703-827-5541
E-mail: clive@idc.com

Name: David Silverman
Duty Title: Chief Scientist
Organization: Innodata
Street: 95 Rockwell Place
City: Brooklyn, NY 11217
Phone: 718-260-4214
E-mail: agman@inod.com

Name: Bryan Layton
Duty Title: Senior Software Consultant
Organization: Intergraph Corp
Street:
City: Huntsville, AL 35894
Phone: 205-730-3069
E-mail: bklayton@ingr.com

Name: Doug Welling
Duty Title: VP, Chief Information Officer
Organization: JANA Inc
Street: 500 N. Kimball Ave #104
City: Southrake, TX 76292
Phone: 817-481-0133
E-mail: dougwelling@msn.com

Name: Holly Smith
Duty Title: Director of Business Development
Organization: Lexicon Systems, Inc
Street: 6165 Lehman, Dr. Ste 204
City: Colorado Springs, CO 80918
Phone: 719-593-8971
E-mail: hollyd@lexisys.com

Name: Chet Ensign
Duty Title: Director
Organization: Logical Design Solutions, Inc
Street: 465 South St. #103
City: Morristown, NJ 07960
Phone: 201-971-0100
E-mail: censign@lds.com

Name: Linda Wolpert
Duty Title: Publications Mgr
Organization: Lucent Technologies
Street: 11900 N Pecos St
City: Denver, CO 80234
Phone: 303-538-4420
E-mail: wolpert@longs.dr.att.com

Name: Roland Brooks
Duty Title: Programmer
Organization: Raytheon Service Co
Street: 2 Wayside Rd
City: Burlington, MA 01803
Phone: 617-238-3096
E-mail: roland@patriots.rsc.ray.com

Name: Janice M. McNulty
Duty Title: Contractor to Solar Turbin, Inc
Organization: SGML Solutions
Street: 2615 Vancouver Ave
City: San Diego, CA 92104
Phone: 619-284-7500
E-mail: mcnultyj@concgntric.ngt

Name: Edward T. Kell
Duty Title: New Product Manager
Organization: Sikorsky Aircraft, Inc.
Street: 6900 Main St
City: Stratford, CT 06497
Phone: 203-384-7289
E-mail: ted-kell@sikorsky.com

Name: Dennis J. OConnor
Duty Title: SGML/DTD Manager
Organization: The Bureau Of National Affairs, Inc
Street: 1231 25th St. NW
City: Washington D.C. 20037
Phone: 202-452-4658
E-mail: doconnor@ban.com

Appendix E: Results of SGML Questionnaire

This questionnaire was sent to 65 commercial enterprises as an exploratory instrument designed to provide information about the organization and its implementation and use of SGML. 20 SGML users provided responses on their use of SGML. The information gathered from this questionnaire will be used to predict the impact of the conversion of Air Force publications/regulations to SGML and how the Air Force can exploit SGML's capabilities. The results of the questionnaire are below. Each category is preceded by the percentage which corresponds to the organizations' responses. These figures do not indicate a level of importance for any one reply over another, rather they indicate the level of frequency in which the organizations replied to the question.

---

1. What is the primary activity of your organization?

30% - Business/industry	20% - Computing	10% - Government	15% - Education
0% - Research	25% - Publishing	5% - Communications	35% - Services
0% - Public Admin	15% - Manufacturing	5% - Retail	0% - Wholesale

20% - Other
   - -- SGML Training and Consulting
   - -- Information Design and Delivery
   - -- Software Development Company
   - -- Standards Writing/Production Certification & Testing
   - -- Systems Integration Consulting

2. How many employees are in your organization?

10% - 0 to 25	30% - 26 to 250	5% - 251 to 500
5% - 501 to 1000	50% - More than 1000	

3. How long has your organization been using SGML?

5% - Less than 1 year	5% - 1 to 2 years	20% - 2 to 3 years
30% - 3 to 5 years	30% - 5 to 7 years	10% - More than 7 years

4. How many people in your organization are using SGML?

   Replies range from 5 to 1750. Mean is 131 people.

5. How many outsourcing contractors in your organization are using SGML?

   Replies range from 0 to 15. Mean is 2.5 contractors.

6. Why have you chosen to use SGML?

5% - Required by upper management
30% - Compliance to a standard or contractual agreement
30% - Other
- -- We are a sponsor of SGML
- -- Required input for browser/delivery platform
- -- Inter and Intranet publishing HTML
- -- Data longevity and on-line distribution /CD-ROM distribution
- -- Single format converted to multiple delivery formats
- -- Cost savings, protect investment in source

60% - Required to meet demand of customers
10% - Competition
50% - Change in information technology

7. How many pages have you converted to SGML?

0% - None
0% - 1 to 100
5% - 101 to 500
0% - 501 to 1000
5% - 1001 to 2000
90% - More than 2000

8. How many SGML documents do you author per year?

5% - None
5% - 1 to 10
20% - 11 to 24
65% - 25 or more

9. For which of the following do you use SGML?

75% - On-line distribution
35% - Document management
90% - Electronic books
55% - Legacy documents
65% - HTML/World Wide Web
35% - Varied platform use
15% - Other
- -- Source from which hardcopy is made
- -- We are a service bureau that creates SGML for customers with all these needs
- -- Parts list database retrieval for IPL's

25% - Storage
60% - Intent for World Wide Web use
65% - CD-ROM
55% - Information preservation
60% - Ensuring structural consistency

10. For what types of documents are you using SGML?

10% - Letters/Memos	25% - Books/Publishing	40% - Electronic Books
65% - User Manuals	25% - Regulations/Publications	35% - On-line Database
90% - Technical Manuals	5% - Request for Proposals	

15% - Other
- -- We produce reference /news type pubs in both print and various electronic formats
- -- Service bulletins, in-house technical letters, integrated parts lists (IPLs)
- -- Software support database

11. What types of information are contained in your documents?

100% - Standard text	95% - Tabular information	100% - Figures and graphics
60% - Math and equations	80% - Links and cross references	

5% - Other
- -- Hyperlinks

12. Are your source documents 60% - Hardcopy (paper) or 65% - Electronic?

13. What percent of authoring is done with the following methods? (based on mean from replies received)

70% - Direct authoring	18% - Generating from a database or spreadsheet
15% - Keyboarding from a manuscript	48% - Scanning from a printed source using OCR
34% - Translating from an unstructured electronic source	0% - Other

14. What operating systems do you use?

55% - MS-DOS	0% - IBM CMS/MVS	25% - Macintosh
70% - UNIX	80% - MS-Windows	0% - OS/2

15% - Other
- -- VMS
- -- MS-Windows NT

15. Which of the following products/tools do you using for conversion/authoring?

40% - FastTag
5% - Intellitag
0% - HyMinder
5% - Write-It
45% - DynaText
0% - Application Builder
45% - In-house applications
-- Various C and Basic and Omnimark programs for conversion
-- Brief and multi-edit editor macros
-- Hot Metal
-- Perl scripts for conversion of parts lists
-- Proteus
-- DCL Tools
60% - Other
-- Adept Editor/Publisher by Arbortext, Inc.
-- Omnimark
-- SGML
-- Framemaker
-- Xychange
-- Trans-SGML
-- Note Pad
-- Passages
-- Near and Far Author
-- WordPerfect SGML
-- Info Access GUIDE Reader

15% - InContext
20% - SGML Hammer
10% - FrameBuilder
0% - SGML Tagger
0% - WorldView

5% - MS Author
0% - SGML/Search
15% - Interleaf
50% - Author/Editor
0% - IBM SGML Translator

16. What do you believe are the benefits of using SGML?

60% - Increased productivity
45% - Better data control
95% - Portability of information
40% - Cost efficiency
50% - Information longevity
5% - Security
15% - Other
-- Use text as a database
-- Multiple outputs, consistent structure and presentation

85% - Reusability
60% - Shareability
55% - Flexibility
35% - Up-to-date information
55% - Improved data integrity
35% - Ability to participate in global markets

17. Has your organization conducted a cost/benefit analysis of SGML in general?

25% - Yes 70% - No

If yes, what were the results:
-- SGML can be justified after initial start-up investment for conversion, s/w acquisition and training. Data re-use is a major cost-saving benefit.
-- Not the worth the effort at present. Need to create highly specified DTD to make implementation worthwhile.
-- Legacy data to SGML is a very good niche for us
-- We have found that authoring in SGML allows us to migrate our documents to a wide range of output specifications, such as HTML, hardcopy, postscript, etc.
-- Cost savings ~ 30% (this savings comes after 10 years of SGML-based publishing)

18. Which of the following have been problems you've encountered using SGML?

40% - Conversion 20% - Authoring 25% - Tagging of documents
40% Training
20% - Other
-- Page composition
-- Explaining benefits of SGML to upper management

Please explain why:
-- Conversion is expensive, unless the data can be re-keyed off-shore, or unless the data already is well structured and consistent
-- Writer buy-in is always a critical link in converting from one system to another
-- Resistance of editors to new ways of marking up data
-- Conversion: because the source was unstructured - manual cleanup was required
-- Conversion of SGML from one form (DTD) to another can be difficult when the target DTD is philosophically dissimilar to the source DTD
-- Lack of experienced people/outsourcing contractors
-- The training is best given by users of the language an not by instructors. This is often the problem
-- Requirement for SGML data to support automatic page composition, imposition and database storage
-- Management doesn't tend to understand document management or document production the way it's been done the past decade; they have a real difficult time with SGML
-- Most writers want WYSIWYG - conversion and authoring (tagging) require new mindset
-- Our DTD is presently too large - adapted directly from the ISO DTD and has 1100 elements of which we regularly use 100.
-- Some authors want WYSIWYG and want immediate style control. Some element relationships are complicated and difficult to master.

19. Does your organization provide SGML training to its employees?

60% - Yes. If so, do you feel this training has enhanced employees' use of SGML?
100% - Yes 0 % - No

35% - No. If not, do you believe training would be beneficial?
85% - Yes 0% - No

20. Does your organization have a training plan for conversion/authoring SGML documents? (For example: Basic SGML training for new employees, in-depth training after 3 months on the job, continuing training after 6 months on the job)

45% - Yes

If yes, how effective is this plan?

44% - Very effective 55% - Meets needs of organization 0% - Not effective

50% - No

If not, would a training plan be beneficial for your organization?

80% - Yes 10% - No 10% - Not Sure

21. What resources have you used to learn about SGML?

80% - Practice/Experience
65% - Co-workers
70% - Training
30% - Other
80% - Books
50% - User groups
40% - Newsletters
40% - Internet

-- We are a teacher/trainer of SGML.
-- SGML 95' Conference and Usergroup meetings
-- We provide corporate SGML training seminars
-- Vendors
-- Consultants

22. I may wish to contact you for further information regarding this questionnaire or SGML. May I call you?

100% - Yes 0% - No

Additional information provided?

- We are totally committed to SGML. It has enabled us to respond quickly to new technologies, such as internet/intranet publishing.
- We still struggle with the conversion of tables and equations. We adopted the CALS standard for this work, but are preparing for an in-depth review of our system and how we work after 4 years of use. We know we are not using SGML to best advantage and require a full review of our DTD.
- SGML is portable, it has an infinite life span, its is totally non-proprietary, and it is a very powerful language. Linked with today's systems with their complex search engines, SGML opens up a whole new era of information technology.
- We are currently working with Adobe to create a custom application using FrameMaker and SGML. We are working with Miles 3 3 to develop a page composition system for fully automated pre-press production. We use Arbortext Adept Editor. We are evaluating database and work flow management systems.
- SGML is flexible in terms of it being processable, but if hardcopy is generated automatically and you need to squeeze one more work on a page, it can be impossible. I have yet to find an authoring tool which is robust and flexible and uncumbersome. That is why we use plain ASCII editors with our own macro enhancements instead of COTS SGML-specific editors.
- We are concerned primarily with electronic distribution of technical information. 4-5 years ago the best tool for this was DynaText; therefore, we adopted SGML and created a very general-purpose DTD. Today, Adobe Acrobat is proving far more efficient and simpler to use. We will therefore cut back on SGML efforts. In the future, we plan to develop a very specific DTD for a subset of our technical information and tag elements by function to facilitate user searches and our own data management. We'll use Frame Builder to do this. The tools to make SGML really work are just not available yet. And the overhead to implement SGML are enormous.

If you have any questions you may contact me at the following address:

Shari T. Miles, Capt, USAF
Graduate Student, Information Resources Management
AFIT/LAR, 2950 P Street
Wright-Patterson AFB, OH 45433-7765

Internet: smiles@afit.af.mil

Fax: 513-476-7988

Bibliography

Alschuler, Liora. "Special Section: Standard General Markup Language," Technical Communications, 40: 208-209 (Second Quarter 1993).

Alschuler, Liora. "Special Section: Standard General Markup Language," Technical Communications, 40: 376-378 (Third Quarter 1993).

Altany, David. "Copycats," Quality Digest, 52-59 (March 1991).

Barron, David. "Why Use SGML?" Electronic Publishing, 2: 3-24 (April 1989).

Brooks, G. "Changes in the Documentation Engineering Process," Electrical Communication, 173-180 (Second Quarter 1994).

Camp, Robert C. Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance. Milwaukee WI: Quality Control, 1989.

Camp, Robert C. "Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance," Quality Progress, 62-83 (January 1989).

Cooper, Donald R. and C. William Emory. Business Research Methods. Chicago: Richard D. Irwin, Inc., 1995.

CQ. Congressional Quarterly Weekly Report, Vol. 50: 3184-3189 (10 October 1992).

Davidson, W.J. "SGML Authoring Tools for Technical Communication," Technical Communication, 403-409 (Third Quarter 1993).

Department of the Air Force. Electronic Publishing Master Program Guide. Washington: HQ USAF, 1996.

Department of the Air Force. The Air Force Publications and Forms Management Programs -- Developing and Processing Publications. AFI 37-160, Volume 1. Washington: HQ USAF, 29 April 1994.

Department of Defense. "DoD Active Duty Military Personnel Strength Levels." Washington: Directorate for Information Operations and Reports, Manpower Management Information Division, 22 June 1996.

Department of Defense. Guide for Assessing Component Information Resources Management Activities. Washington: Office of the Secretary of Defense, 7 September 1995.

Elliot, Ernest (Rear Admiral, US Navy). "CALS: What Should DoD Be Doing?" CALS Journal, 47: 50 (Winter 1993).

Ensign, Chet. "SGML By Evolution," Technical Communications, 40: 387-393 (Third Quarter 1993).

Fawcett, Heather. "The New Oxford English Dictionary Project," Technical Communications, 40: 379-382 (Third Quarter 1993).

Freese, Eric D. "The Transformation of SGML Documents for Presentation on the World Wide Web." WWW, http://141.142.3.76/SDG/IT94/Proceedings/Autools/freese.html (October 1995).

"Getting Started with Online Documentation," Industry Consortium White Paper 4002-II, 1995. WWW, http://www.sgmlopen.org/sgml/docs/abstract.htm.

"Getting Started with SGML." WWW, http://www.sil.org/sgml/arbor2.html (October 1995).

Gilmore, Elizabeth. "Introducing Today's SGML," Technical Communications, 40: 210-217 (Second Quarter 1993).

Glushko, Robert J. and Ken Kershner. "Silicon Graphics' IRIS InSight: An SGML Success Story," Technical Communications, 40: 394-402 (Third Quarter 1993).

Greengard, Samuel. "Does Your HR Department Measure Up? Discover Best Practices Through Benchmarking," Personnel Journal, 74: 63-73 (November 1995).

Gross, Mark. "Getting Your Data Into SGML," Technical Communications, 40: 219-225 (Second Quarter 1993).

Hammer, Michael and James Champy. Reengineering the Corporation: A Manifesto for Business Revolution. New York: Harper Collins Publishers, Inc., 1993.

Harmon, Marion. "Benchmarking," Quality Digest, 20-31 (July 1992).

Heimburger, Anneli. "Introduction to Standard Generalized Markup Language (SGML)," Microcomputers for Information Management, 11: 239-260 (1994).

Lauden, Kenneth C. and Jane Price Lauden. Management Information Systems: Organization and Technology. New Jersey: Prentice-Hall, Inc., 1996.

Lincoln, Sarah and Art Price. "What Benchmarking Books Don't Tell You," Quality Progress, 29: 33-36 (March 1996).

Locke, Christopher. "Making Knowledge Pay," Byte, 17: 245-252 (June 1992).

Lunemann, Rhonda S. "Nuts and Bolts of SGML," Technical Communication, 344-345 (Second Quarter, 1995).

Maunder, Colin. "Documentation on Tap," IEEE Spectrum, 31: 52-56 (September 1994).

Patton, Michael Q. Qualitative Evaluation and Research Methods. Newbury Park CA: SAGE Publications, Inc., 1990.

Port, Otis and Geoffrey Smith. "Beg, Borrow - and Benchmark," Business Week, 74-75 (November 1992).

Rath, Hans Holger and Hans-Peter Wiedling. "Making SGML Work," Computer Standards and Interfaces, 18: 37-53 (January 1996).

Reynolds, Louis R. and Steven J. Derose. "Electronic Books," Byte, 17: 263-268 (June 1992).

Rockley, Ann. "Ontario Hydro and SGML," Technical Communications, 40: 383-386 (Third Quarter 1993).

Salomon, Gary H. "The Use of Intelligent Data for Performance Based Acquisition." WWW, http://www.monmouth.army.mil/cecom/lrc/specstd/int-data.html (17 June 1996).

Schettini, Stephen and Liora Alschuler. "SGML Is Here to Stay," Publish, 9: 71-78 (June 1994).

Schwartau, Winn. Information Warfare: Chaos on the Electronic Superhighway. New York: Thunder's Mouth Press, 1994.

"The SGML Guide." WWW, http://www.sil.org/sgml/ileafgd.html (November 1995).

Smith, Joan M. SGML and Related Standards: Document Description and Processing Languages. New York: Ellis Horwood Limited, 1992.

Stern, David. "SGML Documents: A Better System for Communicating Knowledge," Special Libraries, 86: 117-124 (Spring 1995).

Turner, Linda. "What Every CIO Should Know About SGML," CALS/Enterprise Integration Journal, 53-57 (Summer 1994).

Widnall, Sheila., Secretary of the Air Force. "Air Force Posture." Address to the House National Security Committee. 22 February 1995.

Wood, Eileen M. "So What is SGML?" Journal of Systems Management, 24-29 (September/October 1995).

Wright, Haviland. "SGML Frees Information," Byte, 17: 279-286 (June 1992).

Vita

Captain Shari T. Miles [redacted], [redacted] of Edward and Patricia Bernreuter. She received an Associates of Arts Degree in Data Processing Technology in 1983 from Belleville Area College. In 1984 she enlisted in the U.S. Air Force. Upon graduation from Basic Training, she was assigned to Wilford Hall Medical Center, Lackland AFB, TX, as a Medical Administrative Specialist.

In August 1986, she was accepted into the Reserve Officer Training Corps program at Southwest Texas State University, San Marcos, TX. Captain Miles graduated with a Bachelor of Arts Degree in Occupational Education with an emphasis in Computer Science in May 1988. As a Distinguished Graduate, she received her USAF commission in May 1988 as a Regular Officer.

Captain Miles entered active duty in February 1989 and attended the Communications-Computer Systems Programming and Analysis Officers course at Keesler AFB, MS. Upon graduation she was assigned to the 1st Space Wing, Peterson AFB, CO. In July 1992, Captain Miles was assigned to the 374th Communications Squadron, Yokota AB, Japan. In April 1994, she was assigned as the Executive Officer for the 374th Support Group, Yokota AB. In May 1995, she entered the AFIT School of Logistics and Acquisition Management at Wright-Patterson AFB, OH. She graduated in December 1996 with a Masters degree in Information Resource Management. She was subsequently assigned to Headquarters PACAF, Hickam AFB, HI.

Permanent Address: [redacted]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1996	Master's Thesis

4. TITLE AND SUBTITLE
CONVERTING PUBLICATIONS IN THE AIR FORCE TO SGML: A RECOMMENDATION BASED ON COMMERCIAL EXPERIENCE

5. FUNDING NUMBERS

6. AUTHOR(S)
Shari T. Miles, Captain, USAF

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
2750 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GIR/LAR/96D-5

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Lt Col A. Sgroi, HQ ACC/IMP
230 E. Flightline Rd
Langley AFB, VA 23665-2781

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 Words*)

The objective of this research was to determine those benefits of SGML that the Air Force can use to enhance future publications capabilities and to determine if benchmarking can improve or streamline the Air Force's migration to an electronic publishing environment. To accomplish this, the author completed an exploratory, qualitative study which defined the components and benefits of SGML and the benchmarking application and process. The qualitative techniques for this research included a literature review and questionnaire. The literature review covered both SGML and benchmarking. The questionnaire was an exploratory instrument that was sent to 60 commercial enterprises.

The results of the questionnaire confirmed and elaborated the findings of the literature review. In general, this analysis indicates that the Air Force electronic publishing environment can benefit from the use of SGML.

It was also determined that there are many capabilities and benefits associated with the use of SGML and there are ways in which the Air Force publishing centers can use the full potential of SGML. Last, the analysis indicated that a benchmark of commercial enterprises would provide useful information in streamlining the Air Force's conversion to SGML.

14. SUBJECT TERMS	15. NUMBER OF PAGES
SGML, Benchmarking, Air Force Publications, Information Management	135
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

AFIT Control Number AFIT/GIR/LAR/96D-5

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. **Please return completed questionnaire** to: AIR FORCE INSTITUTE OF TECHNOLOGY/LAC, 2950 P STREET, WRIGHT-PATTERSON AFB OH 45433-7765. Your response is **important.** Thank you.

1. Did this research contribute to a current research project? a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?
a. Yes b. No

3. **Please estimate** what this research would have cost in terms of manpower and dollars if it had been accomplished under contract or if it had been done in-house.

Man Years_____ $_____

4. Whether or not you were able to establish an equivalent value for this research (in Question 3), what is your estimate of its significance?

a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments (Please feel free to use a separate sheet for more detailed answers and include it with this form):

_____ _____
Name and Grade Organization

_____ _____
Position or Title Address